Europäisches Patentamt
European Patent Office
Office européen des brevets

Bescheinigung Certificate Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

02010332.1

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Anmeldung Nr:
Application no.: 02010332.1
Demande no:

Anmeldetag:
Date of filing: 07.05.02
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

DeveloGen Aktiengesellschaft für
entwicklungsbiologische Forschung
Rudolf-Wissell-Strasse 28
37079 Göttingen
ALLEMAGNE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings homologous proteins involved in the regulation of energy homeostasis

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

A61K38/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR

WEICKMANN & WEICKMANN

Patentanwälte
European Patent Attorneys · European Trademark Attorneys

DIPL.-ING. H. WEICKMANN (bis 31.1.01)
DIPL.-ING. F. A. WEICKMANN
DIPL.-CHEM. B. HUBER
DR.-ING. H. LISKA
DIPL.-PHYS. DR. J. PRECHTEL
DIPL.-CHEM. DR. B. BÖHM
DIPL.-CHEM. DR. W. WEISS
DIPL.-PHYS. DR. J. TIESMEYER
DIPL.-PHYS. DR. M. HERZOG
DIPL.-PHYS. B. RUTTENSPERGER
DIPL.-PHYS. DR.-ING. V. JORDAN
DIPL.-CHEM. DR. M. DEY
DIPL.-FORSTW. DR. J. LACHNIT

Unser Zeichen:
28074P EP/WWcl

Anmelder:
DeveloGen Aktiengesellschaft für
entwicklungsbiologische Forschung
Rudolf-Wissell-Straße 28

37079 Göttingen
DEUTSCHLAND

Thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings homolgous proteins involved in the regulation of energy homeostasis

---

Postfach 860 820, 81635 München, Deutschland, Tel. (089) 45563 0, Fax (089) 45563 999, email@weickmann.de

Thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings homolgous proteins involved in the regulation of energy homeostasis

Description

This invention relates to the use of nucleic acid sequences encoding thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings homologous proteins, and the polypeptides encoded thereby and to the use thereof in the diagnosis, study, prevention, and treatment of diseases and disorders related to body-weight regulation, for example, but not limited to, metabolic diseases such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apnea.

Obesity is one of the most prevalent metabolic disorders in the world. It is still poorly understood human disease that becomes more and more relevant for western society. Obesity is defined as an excess of body fat, frequently resulting in a significant impairment of health. Besides severe risks of illness such as diabetes, hypertension and heart disease, individuals suffering from obesity are often isolated socially. Human obesity is strongly influenced by environmental and genetic factors, whereby the environmental influence is often a hurdle for the identification of (human) obesity genes. Obesity is influenced by genetic, metabolic, biochemical, psychological, and behavioral factors. As such, it is a complex disorder that must be addressed on several fronts to achieve lasting positive clinical outcome. Obese individuals are prone to ailments including: diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia,

dyslipidemia, osteoarthritis, gallstones, cancers of the reproductive organs, and sleep apnea.

Obesity is not to be considered as a single disorder but a heterogeneous group of conditions with (potential) multiple causes. Obesity is also characterized by elevated fasting plasma insulin and an exaggerated insulin response to oral glucose intake (Koltermann, J. Clin. Invest 65, 1980, 1272-1284) and a clear involvement of obesity in type 2 diabetes mellitus can be confirmed (Kopelman, Nature 404, 2000, 635-643).

Even if several candidate genes have been described which are supposed to influence the homeostatic system(s) that regulate body mass/weight, like leptin, VCPI, VCPL, or the peroxisome proliferator-activated receptor-gamma co-activator, the distinct molecular mechanisms and/or molecules influencing obesity or body weight/body mass regulations are not known.

Therefore, the technical problem underlying the present invention was to provide for means and methods for modulating (pathological) metabolic conditions influencing body-weight regulation and/or energy homeostatic circuits. The solution to said technical problem is achieved by providing the embodiments characterized in the claims.

Accordingly, the present invention relates to genes with novel functions in body-weight regulation, energy homeostasis, metabolism, and obesity. The present invention discloses specific genes involved in the regulation of body-weight, energy homeostasis, metabolism, and obesity, and thus in disorders related thereto such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apnea. The present invention describes the human thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or

held out wings homologous genes as being involved in those conditions mentioned above.

The term 'GenBank Accession number' relates to NCBI GenBank database entries (Benson et al, Nucleic Acids Res. 28, 2000, 15-18).

Incomplete reduction of atmospheric oxygen generates potent oxidizing agents, including reactive oxygen species (ROS) and their toxic byproducts. Protection from ROS is mediated by nonenzymatic agents, enzymes, and low molecular weight reducing agents, such as thioredoxin. Under normal conditions, thioredoxin reductase reduces oxidized thioredoxin in the presence of NADPH. Reduced thioredoxin serves as an electron donor for thioredoxin peroxidase (peroxiredoxin) which consequently reduces H_2O_2 to H_2O (Schallreuter K.U. and Wood J.M., 2001, J Photochem Photobiol B 64(2-3):179-184). Members of the peroxiredoxin family play an antioxidant protective role in various tissues under nonpathologic conditions and during inflammatory processes. Antioxidants govern intracellular reduction-oxidation (redox) status, which plays a critical role in NFKB (nuclear factror kappa-B) transcription factor activation. Different antioxidants are selective for redox regulation of certain transcription factors. Peroxidases of the peroxiredoxin family reduce hydrogen peroxide (H_2O_2) and alkyl hydroperoxides to water and alcohol with the use of reducing equivalents derived from thiol-containing donor molecules.

A family of highly conserved antioxidant enzymes, Peroxiredoxins (Prxs), has two major Prx subfamilies: one subfamily uses two conserved cysteines (2-Cys) and the other uses 1-Cys to scavenge reactive oxygen species (ROS). Four mammalian 2-Cys members (Prx I-IV) utilize thioredoxin as the electron donor for antioxidation. Prxs are capable of protecting cells from ROS insult and regulating the signal transduction pathways that utilize c-Abl, caspases, nuclear factor-kappaB (NF-kappaB)

and activator protein-1 (AP-1) to influence cell growth and apoptosis. Prxs are also essential for red blood cell (RBC) differentiation and are capable of inhibiting human immunodeficiency virus (HIV) infection and organ transplant rejection (Butterfield L.H. et al., 1999, Antioxid Redox Signal 1(4):385-402). Distribution patterns indicate that Prxs are highly expressed in the tissues and cells at risk for diseases related to ROS toxicity, such as Alzheimer's and Parkinson's diseases and atherosclerosis. This correlation suggests that Prxs are protective against ROS toxicity, yet overwhelmed by oxidative stress in some cells (Butterfield L.H. et al., 1999, Antioxid Redox Signal 1(4):385-402). Prxs tend to form large aggregates at high concentrations, a feature that may interfere with their normal protective function or may even render them cytotoxic. Imbalance in the expression of subtypes can also potentially increase their susceptibility to oxidative stress. Therefor Prxs may play a role in the cellular dysfunction of ROS-related diseases ranging from atherosclerosis to cancer to neurodegenerative diseases.

Cell specificity is based on differential gene expression, which is in part determined by a particular set of transcription factors present and active in a given cell at a certain time. Isoforms of a transcription factor can be expressed at different stages of cell differentiation. Many transcription factors have been identified and characterized, particularly in the liver where there is a wide range of transcriptionally controlled genes. The extinction of many hepatic functions and their reexpression are correlated with the extinction and expression of hepatocyte nuclear factor 4 (HNF4). Moreover, HNF4 has a key role in a transcriptional hierarchy since it also controls the expression of the transcription factor HNF1, which is important in the expression of several hepatic genes. HNF-4 is a steroid hormone receptor super-family member that plays an important role in liver-specific gene expression. The expression pattern of a Drosophila homolog to the mouse transcription factor HNF-4 suggests a determinative role in gut formation (Zhong W. et al., 1993, EMBO J 12(2):537-544).

Hepatocyte nuclear factor (HNF)-4alpha is a transcription factor that plays an important role in regulation of gene expression in pancreatic beta-cells and in the liver. Heterozygous mutations in the HNF-4alpha gene are responsible for maturity-onset diabetes of the young 1 (MODY1), which is characterized by pancreatic beta-cell-deficient insulin secretion. HNF-4alpha is a major transcriptional regulator of many genes expressed in the liver. A heterozygous HNF-4alpha mutation leads to an HNF-4alpha-dependent hepatocyte secretory defect of liver-specific proteins (see, for example, Shih D.Q. et al., 2000, Diabetes 49(5):832-837 or Ryffel G.U., 2001, J Mol Endocrinol 27(1):11-29).

Dietary fat is an important macronutrient for the growth and development of all organisms. Specific fatty acid-regulated transcription factors have been identified in bacteria, amphibians, and mammals. In mammals, these factors include peroxisome proliferator-activated receptors (PPAR alpha, -beta, and -gamma), HNF4 alpha, NF kappa B, and SREBP1c (Jump D.B. and Clarke S.D., 1999, Annu Rev Nutr 19:63-90).

The myelin basic protein (MBP) gene is expressed in oligodendrocytes and Schwann cells, and expression follows a tightly regulated developmental time course. Cell type- and developmental stage-specific expression of the MBP gene is regulated by a series of cis-acting elements located upstream of the transcription start site. Myelin gene expression factor-2 (Myef-2), a protein isolated from mouse brain represses transcription of the MBP gene. Myef-2 mRNA is developmentally regulated in mouse brain; its peak expression occurs at postnatal day 7, prior to the onset of MBP expression (Haas S. et al., 1995, J Biol Chem 270(21):12503-12510).

MBP is a major component of the myelin sheath whose production is developmentally controlled during myelinogenesis. Programmed expression of the MBP gene is regulated at the level of transcription. The MB1 regulatory motif plays an important role in transcription of the MBP

promoter. The MB1 element contains a binding site for the repressor protein MyEF-2 (Myelin gene expression factor-2). MyEF-2 is involved in transcriptional regulation of the MBP gene during the course of brain development (Muralidharan V. et al., 1997, J Cell Biochem 1997 Sep 15;66(4):524-31).

The held out wings (how) Drosophila gene encodes a RNA-binding protein involved in the control of muscular and cardiac activity. The how protein is localized to the nucleus. how is highly related to the mouse quaking gene which plays a role at least in myelination and that could serve to link a signal transduction pathway to the control of mRNA metabolism (Zaffran S. et al., 1997, Development 124(10):2087-2098). Two isoforms of the Drosophila RNA binding protein, how, act in opposing directions to regulate tendon cell differentiation (Nabel-Rosen H. et al., 2002, Dev Cell 2002 Feb;2(2):183-193). The opposing activities of the How isoforms are manifested by differential rates of mRNA degradation of the target stripe mRNA. This mechanism is conserved, as the mammalian RNA binding Quaking proteins may similarly affect the levels of Krox20, a regulator of Schwann cell maturation.

The mouse quaking (qk) gene is essential in both myelination and early embryogenesis. Its product, QKI, is an RNA-binding protein belonging to a growing protein family called STAR (signal transduction and activator of RNA) (Wu J. et al., 1999, J Biol Chem 274(41):29202-29210). Quaking is essential for blood vessel development (Noveroske J.K. et al., 2002, Genesis 32(3):218-230).

So far, it has not been described that thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins are involved in the regulation of energy homeostasis and body-weight regulation and related disorders, and thus, no functions in metabolic diseases and other diseases as listed above have been discussed. In this

invention we demonstrate that the correct gene dose of thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings is essential for maintenance of energy homeostasis. A genetic screen was used to identify that mutation of thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings homologous genes cause obesity, reflected by a significant change of triglyceride content, the major energy storage substance.

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention that will be limited only by the appended claims. Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing the cell lines, vectors, and methodologies that are reported in the publications which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure.

The present invention discloses that thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings homologous proteins are regulating the energy homeostasis and fat metabolism especially the metabolism and storage of triglycerides, and polynucleotides, which identify and encode the proteins disclosed in this invention. The invention

also relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides of the invention. The invention also relates to the use of these sequences in the diagnosis, study, prevention, and treatment of diseases and disorders, for example, but not limited to, metabolic diseases such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apnea.

Thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings homologous proteins and nucleic acid molecules coding therefore are obtainable from insect or vertebrate species, e.g. mammals or birds. Particularly preferred are homologous nucleic acids, particularly nucleic acids encoding the human peroxiredoxin family, the human hepatocyte nuclear factor 4 family, human myelin expression factor 2 and closely related proteins, and the human quaking isoforms.

The invention particularly relates to a nucleic acid molecule encoding a polypeptide contributing to regulating the energy homeostasis and the metabolism of triglycerides, wherein said nucleic acid molecule comprises

(a) the nucleotide sequence of (i) Jafrac1 (GadFly Accession Number CG1633), a nucleic acid of the human peroxiredoxin family, (ii) Hnf4 (GadFly Accession Number CG9310), a nucleic acid of the human hepatocyte nuclear factor 4 family, (iii) CG9373 (GadFly Accession Number), a nucleic acid of human myelin expression factor 2 gene and closely related genes, or (iv) how (GadFly Accession Number CG10293), or a nucleic acid of the human quaking isoform genes, and/or a sequence complementary thereto,

(b) a nucleotide sequence which hybridizes at 50°C in a solution containing 1 x SSC and 0.1% SDS to a sequence of (a),

(c) a sequence corresponding to the sequences of (a) or (b) within the degeneration of the genetic code,

(d) a sequence which encodes a polypeptide which is at least 85%, preferably at least 90%, more preferably at least 95%, more preferably at least 98% and up to 99,6% identical to the amino acid sequences of thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings protein, preferably of a protein of the human peroxiredoxin family, a protein of the human hepatocyte nuclear factor 4 family, human myelin expression factor 2 and closely related proteins, or the human quaking isoforms,

(e) a sequence which differs from the nucleic acid molecule of (a) to (d) by mutation and wherein said mutation causes an alteration, deletion, duplication and/or premature stop in the encoded polypeptide or

(f) a partial sequence of any of the nucleotide sequences of (a) to (e) having a length of at least 15 bases, preferably at least 20 bases, more preferably at least 25 bases and most preferably at least 50 bases.

The present invention relates to genes with novel functions in body-weight regulation, energy homeostasis, metabolism, and obesity. The invention is based on the finding that thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings homologous proteins (herein referred to as thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings) and the polynucleotides encoding these, are involved in the regulation of triglyceride storage and therefore energy homeostasis. To find genes with novel functions in energy homeostasis, metabolism, and obesity, a functional genetic screen was performed with the model organism Drosophila melanogaster (Meigen). One resource for screening was a Drosophila melanogaster stock collection of EP-lines. The P-vector of this collection has Gal4-UAS-binding sites fused to a basal promoter that can transcribe adjacent genomic Drosophila sequences upon binding of

Gal4 to UAS-sites. This enables the EP-line collection for overexpression of endogenous flanking gene sequences. In addition, without activation of the UAS-sites, integration of the EP-element into the gene is likely to cause a reduction of gene activity, and allows determining its function by evaluating the loss-of-function phenotype.

Obese people mainly show a significant increase in the content of triglycerides. Triglycerides are the most efficient storage for energy in cells. In order to isolate genes with a function in energy homeostasis, several thousand EP-lines were tested for their triglyceride content after a prolonged feeding period (see Examples for more detail). Lines with significantly changed triglyceride content were selected as positive candidates for further analysis. The increase or decrease of triglyceride content due to the loss of a gene function suggests gene activities in energy homeostasis in a dose dependent manner that controls the amount of energy stored as triglycerides.

In this invention, the content of triglycerides of a pool of flies with the same genotype after feeding for six days was analyzed using a triglyceride assay. Male flies homozygous for the integration of vectors for Drosophila lines PX9430.2, EP(2)2449, HD-EP(3)31646, or HD-EP(3)30815 were analyzed in an assay measuring the triglyceride contents of these flies, illustrated in more detail in the EXAMPLES section. The results of the triglyceride content analysis are shown in FIGURES 1, 4, 7, and 10, respectively.

Genomic DNA sequences were isolated that are localized to the EP vector (herein PX9430.2, EP(2)2449, HD-EP(3)31646, or HD-EP(3)30815) integration. Using those isolated genomic sequences public databases like Berkeley Drosophila Genome Project (GadFly) were screened thereby identifying the integration site of the vectors, and the corresponding genes, described in more detail in the EXAMPLES section. The molecular

organization of the genes is shown in FIGURES 2, 5, 8, and 11, respectively.

The present invention further describes polypeptides comprising the amino acid sequences of thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins. Based upon homology, the proteins of the invention and each homologous protein or peptide may share at least some activity. No functional data described the regulation of body weight control and related metabolic diseases are available in the prior art for the genes of the invention.

The invention also encompasses polynucleotides that encode thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins. Accordingly, any nucleic acid sequence, which encodes the amino acid sequences of thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins, can be used to generate recombinant molecules that express thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins. In a particular embodiment, the invention encompasses a nucleic acid encoding Jafrac1 (GadFly Accession Number CG1633), a protein of the human peroxiredoxin family, Hnf4 (GadFly Accession Number CG9310), a protein of the human hepatocyte nuclear factor 4 family, CG9373 (GadFly Accession Number), human myelin expression factor 2 and closely related proteins, how (GadFly Accession Number CG10293), or the human quaking isoforms. It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding the proteins, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet

genetic code as applied to the nucleotide sequences of naturally occurring thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins, and all such variations are to be considered as being specifically disclosed. Although nucleotide sequences, which encode the proteins, and their variants are preferably capable of hybridizing to the nucleotide sequences of the naturally occurring proteins under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding the proteins or their derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding the proteins and their derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequences. The invention also encompasses production of DNA sequences, or portions thereof, which encode the proteins and their derivatives, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art at the time of the filing of this application. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding the protein or any portion thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed nucleotide sequences, and in particular, those of the polynucleotide encoding Jafrac1 (GadFly Accession Number CG1633), a protein of the human peroxiredoxin family, Hnf4 (GadFly Accession Number CG9310), a protein of the human hepatocyte nuclear factor 4 family, CG9373 (GadFly Accession Number), human myelin expression factor 2 and closely related proteins, how (GadFly

Accession Number CG10293), or the human quaking isoforms, under various conditions of stringency. Hybridization conditions are based on the melting temperature (Tm) of the nucleic acid binding complex or probe, as taught in Wahl, G. M. and S. L. Berger (1987: Methods Enzymol. 152:399-407) and Kimmel, A. R. (1987; Methods Enzymol. 152:507-511), and may be used at a defined stringency. Preferably, hybridization under stringent conditions means that after washing for 1 h with 1 x SSC and 0.1% SDS at 50°C, preferably at 55°C, more preferably at 62°C and most preferably at 68°C, particularly for 1 h in 0.2 x SSC and 0.1% SDS at 50°C, preferably at 55°C, more preferably at 62°C and most preferably at 68°C, a positive hybridization signal is observed. Altered nucleic acid sequences encoding the proteins which are encompassed by the invention include deletions, insertions, or substitutions of different nucleotides resulting in a polynucleotide that encodes the same or a functionally equivalent protein.

The encoded proteins may also contain deletions, insertions, or substitutions of amino acid residues, which produce a silent change and result in functionally equivalent proteins. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the biological activity of the protein is retained.

Also included within the scope of the present invention are alleles of the genes encoding thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins. As used herein, an 'allele' or 'allelic sequence' is an alternative form of the gene, which may result from at least one mutation in the nucleic acid sequence. Alleles may result in altered mRNAs or polypeptides whose structures or function may or may not be altered. Any given gene may have none, one, or many allelic forms. Common mutational changes, which give rise to alleles, are generally ascribed to natural deletions, additions, or substitutions of

nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence. Methods for DNA sequencing which are well known and generally available in the art may be used to practice any embodiments of the invention.

The nucleic acid sequences encoding thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins may be extended utilizing a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one method which may be employed, 'restriction-site' PCR, uses universal primers to retrieve unknown sequence adjacent to a known locus (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Inverse PCR may also be used to amplify or extend sequences using divergent primers based on a known region (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). The primers may be designed using OLIGO 4.06 primer analysis software (National Biosciences Inc., Plymouth, Minn.), or another appropriate program, to 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68-72°C. The method uses several restriction enzymes to generate suitable fragments. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Another method which may be used is capture PCR which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations also are used to place an engineered double-stranded sequence into an unknown portion of the DNA molecule before performing PCR.

Another method which may be used to retrieve unknown sequences is that of Parker, J. D. et al. (1991; Nucleic Acids Res. 19:3055-3060).

Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries to walk in genomic DNA (Clontech, Palo Alto, Calif.). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable, in that they will contain more sequences, which contain the 5′ regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into the 5′ and 3′ non-transcribed regulatory regions. Capillary electrophoresis systems, which are commercially available, may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and detection of the emitted wavelengths by a charge coupled devise camera. Output/light intensity may be converted to electrical signal using appropriate software (e.g. GENOTYPER and SEQUENCE NAVIGATOR, Perkin Elmer) and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for the sequencing of small pieces of DNA, which might be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins, or fusion proteins or functional equivalents thereof, may be used in recombinant DNA molecules to direct expression of the proteins in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences, which encode substantially the same, or a

functionally equivalent amino acid sequence may be produced and these sequences may be used to clone and express the proteins. As will be understood by those of skill in the art, it may be advantageous to produce protein-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce a recombinant RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence. The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter protein-encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, or introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen libraries, e.g. peptide libraries or low-molecular weight compound libraries for inhibitors of thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous protein activities, it may be useful to encode chimeric proteins that can be recognized by a commercially available antibodies. A fusion protein may also be engineered to contain a cleavage site located between the desired protein-encoding sequence and the heterologous protein sequence so that the desired protein may be cleaved and purified away from the heterologous moiety. In another embodiment, sequences encoding the

protein may be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers, M. H. et al. (1980) Nucl. Acids Res. Symp. Ser. 7:215-223, Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 7:225-232). Alternatively, the proteins themselves may be produced using chemical methods to synthesize the amino acid sequence of the protein, or a portion thereof. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, J. Y. et al. (1995) Science 269:202-204) and automated synthesis may be achieved, for example, using the ABI 431A peptide synthesizer (Perkin Elmer). The newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography (e.g., Creighton, T. (1983) Proteins, Structures and Molecular Principles, WH Freeman and Co., New York, N.Y.). The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure; Creighton, supra). Additionally, the amino acid sequences of the proteins, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a biologically active protein, the nucleotide sequences encoding the proteins or functional equivalents, may be inserted into appropriate expression vectors, i.e., a vector, which contains the necessary elements for the transcription and translation of the inserted coding sequence. Methods, which are well known to those skilled in the art, may be used to construct expression vectors containing sequences encoding the proteins and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. Such techniques are described in Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y., and Ausubel, F. M. et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y.

A variety of expression vector/host systems may be utilized to contain and express sequences encoding the proteins. These include, but are not limited to, micro-organisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or PBR322 plasmids); or animal cell systems. The "control elements" or "regulatory sequences" are those non-translated regions of the vector-enhancers, promoters, 5′ and 3′ untranslated regions which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the BLUESCRIPT phagemid (Stratagene, LaJolla, Calif.) or PSPORT1 plasmid (Gibco BRL) and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters and enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO; and storage protein genes) or from plant viruses (e.g., viral promoters and leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequences encoding the protein, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for the protein. For example, when large quantities of protein are needed for the induction of antibodies, vectors, which direct high level expression of fusion proteins that are readily

purified, may be used. Such vectors include, but are not limited to, the multifunctional E. coli cloning and expression vectors such as the BLUESCRIPT phagemid (Stratagene), in which the sequences encoding the protein may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of ß-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke, G. and S. M. Schuster (1989) J. Biol. Chem. 264:5503-5509); and the like. PGEX vectors (Promega, Madison, Wis.) may also be used to express foreign polypeptides as fusion proteins with Glutathione S-Transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will. In the yeast, Saccharomyces cerevisiae, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, see Ausubel et al., (supra) and Grantet al. (1987) Methods Enzymol. 153:516-544.

In cases where plant expression vectors are used, the expression of sequences encoding the proteins may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, for example, Hobbs, S. or

Murry, L. E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, N.Y.; pp. 191-196).

An insect system may also be used to express the proteins. For example, in one such system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in Spodoptera frugiperda cells or in Trichoplusia larvae. The sequences encoding the protein may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and place under control of the polyhedrin promoter. Successful insertions of the protein will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, S. frugiperda cells of Trichoplusia larvae in which thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins may be expressed (Engelhard, E. K. et al. (1994) Proc. Nat. Acad. Sci. 91:3224-3227).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding the protein may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain viable viruses which are capable of expressing the protein in infected host cells (Logan, J. and Shenk, T. (1984) Proc. Natl. Acad. Sci. 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding the protein. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding the protein, its initiation codons, and upstream sequences are

inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a 'prepro' form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells such as CHO, HeLa, MDCK, HEK293, and WI38, which have specific cellular machinery and characteristic mechanisms for such post-translational activities, may be chosen to ensure the correct modification and processing of the foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins may be generated by transformation using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before

they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells, which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type. Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler, M. et al. (1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. et al. (1980) Cell 22:817-23) genes, which can be employed in tk-or aprt-,cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have been described, for example, trpB, which allows cells to utilise indole in place of tryptophan, or hisD, which allows cells to utilise histinol in place of histidine (Hartman, S. C. and R. C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-51). Recently, the use of visible markers has gained popularity with such markers as anthocyanins, ß- glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C. A. et al. (1995) Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the sequences encoding the protein of interest are inserted within a marker gene sequence, recombinant cells containing sequences encoding the protein can be identified by the

absence of marker gene function. Alternatively, a marker gene can be placed in tandem with sequences encoding the protein under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well. Alternatively, host cells, which contain and express the nucleic acid sequences encoding the protein may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA, or DNA-RNA hybridization and protein bioassay or immunoassay techniques that include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein.

The presence of polynucleotide sequences encoding thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes or portions or fragments of polynucleotides encoding thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences specific for the gene to detect transformants containing DNA or RNA encoding the corresponding protein. As used herein 'oligonucleotides' or 'oligomers' refer to a nucleic acid sequence of at least about 10 nucleotides and as many as about 60 nucleotides, preferably about 15 to 30 nucleotides, and more preferably about 20-25 nucleotides, which can be used as a probe or amplimer.

A variety of protocols for detecting and measuring the expression of proteins, using either polyclonal or monoclonal antibodies specific for the protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on

the protein is preferred, but a competitive binding assay may be employed. These and other assays are described, among other places, in Hampton, R. et al. (1990; Serological Methods, a Laboratory Manual, APS Press, St Paul, Minn.) and Maddox, D. E. et al. (1983; J. Exp. Med. 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins include oligo-labeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide.

Alternatively, the sequences encoding the protein, or any portions thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits (Pharmacia & Upjohn, (Kalamazoo, Mich.); Promega (Madison Wis.); and U.S. Biochemical Corp., (Cleveland, Ohio).

Suitable reporter molecules or labels, which may be used, include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, co-factors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding the protein may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence

and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode the protein may be designed to contain signal sequences, which direct secretion of the protein through a prokaryotic or eukaryotic cell membrane. Other recombinant constructions may be used to join sequences encoding the protein to nucleotide sequence encoding a polypeptide domain, which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAG extension/affinity purification system (Immunex Corp., Seattle, Wash.) The inclusion of cleavable linker sequences such as those specific for Factor XA or Enterokinase (Invitrogen, San Diego, Calif.) between the purification domain and the desired protein may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing the desired protein and a nucleic acid encoding 6 histidine residues preceding a thioredoxine or an Enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilised metal ion affinity chromatography as described in Porath, J. et al. (1992, Prot. Exp. Purif. 3: 263-281)) while the Enterokinase cleavage site provides a means for purifying the desired protein from the fusion protein. A discussion of vectors which are suitable for the production of fusion proteins is provided in Kroll, D. J. et al. (1993; DNA Cell Biol. 12:441-453). In addition to recombinant production, fragments of the proteins may be produced by direct peptide synthesis using solid-phase techniques (Merrifield J. (1963) J. Am. Chem. Soc. 85:2149-2154). Protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A peptide synthesizer (Perkin Elmer). Various fragments of the proteins may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

Diagnostics and Therapeutics

The data disclosed in this invention show that the nucleic acids and proteins of the invention are useful in diagnostic and therapeutic applications implicated, for example but not limited to, in metabolic disorders such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis; gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apnea. Hence, diagnostic and therapeutic uses for the thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings nucleic acids and proteins of the invention are, for example but not limited to, the following: (i) protein therapeutic, (ii) small molecule drug target, (iii) antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) diagnostic and/or prognostic marker, (v) gene therapy (gene delivery/gene ablation), (vi) research tools, and (vii) tissue regeneration in vitro and in vivo (regeneration for all these tissues and cell types composing these tissues and cell types derived from these tissues).

The nucleic acids and proteins of the invention are useful in diagnostic and therapeutic applications implicated in various applications as described below. For example, but not limited to, cDNAs encoding the thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings proteins of the invention and particularly their human homologues may be useful in gene therapy, and the thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings proteins of the invention and particularly their human homologues may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from, for example, but not limited to, in metabolic disorders as described above.

The novel nucleic acid encoding the thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings protein of the invention, or homologous proteins, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acids or the proteins are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

For example, in one aspect, antibodies which are specific for thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins may be used directly as an antagonist, or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express the protein. The antibodies may be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimerical, single chain, Fab fragments, and fragments produced by a Fab expression library. Neutralising antibodies, (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others, may be immunized by injection with the protein or any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminium hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. Among adjuvants used in human, BCG (Bacille Calmette-Guerin) and Corynebacterium parvum are especially preferable. It is preferred that the peptides, fragments, or oligopeptides used to induce antibodies to the protein have an amino acid sequence consisting of at least five amino acids, and more preferably at least 10 amino acids. It is preferable that

they are identical to a portion of the amino acid sequence of the natural protein, and they may contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous protein amino acids may be fused with those of another protein such as keyhole limpet hemocyanin in order to increase the immunogenicity.

Monoclonal antibodies to the proteins may be prepared using any technique that provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Köhler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R. J. et al. Proc. Natl. Acad. Sci. 80:2026-2030; Cole, S. P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of 'chimeric antibodies', the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can be used (Morrison, S. L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M. S. et al (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce single chain antibodies specific for thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D. R. (1991) Proc. Natl. Acad. Sci. 88:11120-3). Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening recombinant immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R.

et al. (1989) Proc. Natl. Acad. Sci. 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for the proteins may also be generated. For example, such fragments include, but are not limited to, the F(ab')2 fragments which can be produced by Pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W. D. et al. (1989) Science 254:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding and immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between the protein and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering protein epitopes are preferred, but a competitive binding assay may also be employed (Maddox, supra).

In another embodiment of the invention, the polynucleotides encoding thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins, or any fragment thereof, or antisense molecules, may be used for therapeutic purposes. In one aspect, antisense molecules may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins. Thus, antisense molecules may be used to modulate protein activity, or to achieve regulation of gene function. Such technology

is now well know in the art, and sense or antisense oligomers or larger fragments, can be designed from various locations along the coding or control regions of sequences encoding the proteins. Expression vectors derived from retroviruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids may be used for delivery of nucleotide sequences to the targeted organ, tissue or cell population. Methods, which are well known to those skilled in the art, can be used to construct recombinant vectors, which will express antisense molecules complementary to the polynucleotides of the genes encoding thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins. These techniques are described both in Sambrook et al. (supra) and in Ausubel et al. (supra). Genes encoding thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins can be turned off by transforming a cell or tissue with expression vectors which express high levels of polynucleotide which encodes thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins or fragments thereof. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector and even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing antisense molecules, e.g. DNA, RNA, or PNA, to the control regions of the genes encoding thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins, i.e., the promoters, enhancers, and introns. Oligonucleotides derived from the transcription initiation site, e.g., between positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful

because it cause inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J. E. et al. (1994) In; Huber, B. E. and B. I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y.). The antisense molecules may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Examples, which may be used, include engineered hammerhead motif ribozyme molecules that can be specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins. Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Antisense molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA

sequences encoding thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins. Such DNA sequences may be incorporated into a variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize antisense RNA constitutively or inducibly can be introduced into cell lines, cells, or tissues. RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5′ and/or 3′ ends of the molecule or the use of phosphorothioate or 2′ O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of non-traditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection and by liposome injections may be achieved using methods, which are well known in the art. Any of the therapeutic methods described above may be applied to any suitable subject including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and

homologous nucleic acids or proteins, antibodies to thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins, mimetics, agonists, antagonists, or inhibitors of thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins or nucleic acids. The compositions may be administered alone or in combination with at least one other agent, such as stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs or hormones. The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries, which facilitate processing of the active compounds into preparations which, can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, Pa.). Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding

suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums including Arabic and tragacanth; and proteins such as gelatine and collagen. If desired, disintegrating or solubilising agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate. Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum Arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coating for product identification or to characterize the quantity of active compound, i.e., dosage. Pharmaceutical preparations, which can be used orally, include push-fit capsules made of gelatine, as well as soft, sealed capsules made of gelatine and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances, which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such

as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Optionally, the suspension may also contain suitable stabilizers or agents who increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes. The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulphuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1-50 mM histidine, 0.1%-2% sucrose, and 2-7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use. After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of proteins, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art. For any compounds, the therapeutically effective does can be estimated initially either in cell culture assays, e.g., of preadipocyte cell lines, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be

used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. A therapeutically effective dose refers to that amount of active ingredient, for example thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins or nucleic acids or fragments thereof, antibodies of thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins,which is sufficient for treating a specific condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50. Pharmaceutical compositions, which exhibit large therapeutic indices, are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage varies within this range depending upon the dosage from employed, sensitivity of the patient, and the route of administration. The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors, which may be taken into account, include the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation. Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the

route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

In another embodiment, antibodies which specifically bind to the proteins may be used for the diagnosis of conditions or diseases characterized by or associated with over- or underexpression of thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins, or in assays to monitor patients being treated with thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins, agonists, antagonists or inhibitors. The antibodies useful for diagnostic purposes may be prepared in the same manner as those described above for therapeutics. Diagnostic assays include methods which utilize the antibody and a label to detect the protein in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules which are known in the art may be used several of which are described above.

A variety of protocols including ELISA, RIA, and FACS for measuring proteins are known in the art and provide a basis for diagnosing altered or abnormal levels of gene expression. Normal or standard values for gene expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibodies to the protein under conditions suitable for complex formation. The amount of standard complex formation may be quantified by various methods, but preferably by photometry, means. Quantities of protein expressed in control and disease, samples from biopsied tissues are compared with the

standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides specific for thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins may be used for diagnostic purposes. The polynucleotides, which may be used, include oligonucleotide sequences, antisense RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which gene expression may be correlated with disease. The diagnostic assay may be used to distinguish between absence, presence, and excess gene expression, and to monitor regulation of protein levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins or closely related molecules, may be used to identify nucleic acid sequences which encode the respective protein. The specificity of the probe, whether it is made from a highly specific region, e.g., unique nucleotides in the 5' regulatory region, or a less specific region, e.g., especially in the 3' coding region, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low) will determine whether the probe identifies only naturally occurring sequences, alleles, or related sequences. Probes may also be used for the detection of related sequences, and should preferably contain at least 50% of the nucleotides from any of the thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous protein-encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and derived from the nucleotide sequence of the polynucleotide comprising Jafrac1 (GadFly Accession Number CG1633), a nucleic acid of the human peroxiredoxin family, Hnf4

(GadFly Accession Number CG9310), a nucleic acid of the human hepatocyte nuclear factor 4 family, CG9373 (GadFly Accession Number), a nucleic acid of the human myelin expression factor 2 gene and closely related genes, how (GadFly Accession Number CG10293), or a nucleic acid of the human quaking isoform genes or from a genomic sequence including promoter, enhancer elements, and introns of the naturally occurring gene. Means for producing specific hybridization probes for DNAs encoding thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins include the cloning of nucleic acid sequences specific for thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins into vectors for the production of mRNA probes. Such vectors are known in the art, commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, radionuclides such as ^{32}P or ^{35}S, or enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences specific for thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous nucleic acids may be used for the diagnosis of conditions or diseases, which are associated with the expression of the proteins. Examples of such conditions or diseases include, but are not limited to, pancreatic diseases and disorders, including diabetes. Polynucleotide sequences specific for thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins may also be used to monitor the progress of patients receiving treatment for pancreatic diseases and disorders, including diabetes. The polynucleotide sequences may be used in Southern or Northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; or in dip stick, pin, ELISA or chip assays utilizing fluids or tissues from patient biopsies to detect altered gene

expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences specific for thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous nucleic acids may be useful in assays that detect activation or induction of various metabolic diseases such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancers of the reproductive organs, and sleep apnea. The nucleotide sequences may be labeled by standard methods, and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the biopsied or extracted sample is significantly altered from that of a comparable have hybridized with nucleotide sequences in the sample, and the presence of altered levels of nucleotide sequences encoding thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins in the sample indicates the presence of the associated disease. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or in monitoring the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disease associated with expression of thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, which is specific for thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous nucleic acids, under conditions suitable for

hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with those from an experiment where a known amount of a substantially purified polynucleotide is used. Standard values obtained from normal samples may be compared with values obtained from samples from patients who are symptomatic for disease. Deviation between standard and subject values is used to establish the presence of disease. Once disease is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to evaluate whether the level of expression in the patient begins to approximate that, which is observed in the normal patient. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to metabolic diseases such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancers of the reproductive organs, and sleep apnea the presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the pancreatic diseases and disorders. Additional diagnostic uses for oligonucleotides designed from the sequences encoding thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins may involve the use of PCR. Such oligomers may be chemically synthesized, generated enzymatically, or produced from a recombinant source. Oligomers will preferably consist of two nucleotide sequences, one with sense orientation (5'.fwdarw.3') and another with antisense (3'.rarw.5'), employed under optimized conditions for identification of a specific gene or condition. The same two oligomers,

nested sets of oligomers, or even a degenerate pool of oligomers may be employed under less stringent conditions for detection and/or quantification of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and standard curves onto which the experimental results are interpolated (Melby, P. C. et al. (1993) J. Immunol. Methods, 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantification of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantification.

In another embodiment of the invention, the nucleic acid sequences which are sprecific for thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous nucleic acids may also be used to generate hybridization probes, which are useful for mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome or to a specific region of the chromosome using well known techniques. Such techniques include FISH, FACS, or artificial chromosome constructions, such as yeast artificial chromosomes, bacterial artificial chromosomes, bacterial P1 constructions or single chromosome cDNA libraries as reviewed in Price, C. M. (1993) Blood Rev. 7:127-134, and Trask, B. J. (1991) Trends Genet. 7:149-154. FISH (as described in Verma et al. (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York, N.Y.) may be correlated with other physical chromosome mapping techniques and genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of the gene encoding thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held

out wings on a physical chromosomal map and a specific disease, or predisposition to a specific disease, may help to delimit the region of DNA associated with that genetic disease.

The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier, or affected individuals. In situ hybridization of chromosomal preparations and physical mapping techniques such as linkage analysis using established chromosomal markers may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms, or parts thereof, by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, for example, AT to 11q22-23 (Gatti, R. A. et al. (1988) Nature 336:577-580), any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleotide sequences of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier, or affected individuals.

In another embodiment of the invention, thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins, their catalytic or immunogenic fragments or oligopeptides thereof, can be used for screening libraries of compounds, e.g. peptides or low-molecular weight organic compounds, in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held

out wings and homologous proteins and the agent tested, may be measured.

Another technique for drug screening, which may be used, provides for high throughput screening of compounds having suitable binding affinity to the protein of interest as described in published PCT application WO84/03564. In this method, as applied to thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous proteins large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with the proteins, or fragments thereof, and washed. Bound proteins are then detected by methods well known in the art. Purified proteins can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support. In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding the protein specifically compete with a test compound for binding the protein. In this manner, the antibodies can be used to detect the presence of any peptide, which shares one or more antigenic determinants with the protein. In additional embodiments, the nucleotide sequences which are specific for thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings and homologous nucleic acids or proteins encoded thereby may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

The Figures show:

FIGURE 1 shows the increase of triglyceride content of PX9430.2 flies ('PX 9430.2', column 2) caused by homozygous viable integration of the

P-vector into the leader of the CG1633 gene (in comparison to controls without integration of this vector; 'PX-control', column 1).

FIGURE 2 shows the molecular organization of the mutated CG1633 (Gadfly Accession Number) gene locus.

FIGURE 3 shows the BLASTP search result for the CG1633 gene product (Query) with the four best human homologous matches (Sbjct).

FIGURE 4 shows the decrease of triglyceride content of EP(2)2449 ('EP(2)2449', column 2) flies caused by homozygous viable integration of the P-vector into an intron of CG9310 (in comparison to controls without integration of this vector, 'PX-control', column 1).

FIGURE 5 shows the molecular organization of the mutated inx2 (Gadfly Accession Number CG9310) gene locus.

FIGURE 6 shows the BLASTP search result for the CG9310 gene product (Query) with the four best human homologous matches (Sbjct).

FIGURE 7 shows the triglyceride content of a CG9373 protein mutant. Shown is the increase of triglyceride content of HD-EP(3)31646 flies ('HD-EP3646/elav', column 2) caused by ectopic expression of the CG9373 gene (in comparison to controls with integration of this vector, 'random EP/elav', column 1) mainly in the neurons of these flies.

FIGURE 8 shows the molecular organization of the mutated CG9373 (Gadfly Accession Number) gene locus.

FIGURE 9 shows the BLASTP search result for the CG9373 gene product (Query) with the three best human homologous matches (Sbjct).

FIGURE 10 shows the decrease of triglyceride content of HD-EP(3)30815 flies ('HD-EP30815', column 2) caused by homozygous viable integration of the P-vector into the promoter of the CG10293 gene (in comparison to controls without integration of this vector, 'EP-control', column 1).

FIGURE 11 shows the molecular organization of the mutated CG10293 (Gadfly Accession Number) gene locus.

FIGURE 12 shows the BLASTP search result for the CG10293 gene product (Query) with the twelve best human homologous matches (Sbjct).

The examples illustrate the invention:

Example 1: Measurement of triglyceride content

Mutant flies are obtained from a fly mutation stock collection. The flies are grown under standard conditions known to those skilled in the art. In the course of the experiment, additional feedings with bakers yeast (Saccharomyces cerevisiae) are provided for the EP-lines EP(2)2449, HD-EP(3)31646, and HD-EP(3)30815. The average increase or decrease of triglyceride content of Drosophila containing the EP-vectors as homozygous viable integration was investigated in comparison to control flies (see FIGURES 1, 4, 7, and 10). For determination of triglyceride, flies (in case of PX9430.2, ten flies in two independent assays, respectively) were incubated for 5 min at 90(C (in case of PX9430.2 at 70°C) in an aqueous buffer using a waterbath, followed by hot extraction. After another 5 min incubation at 90(C (in case of PX9430.2 at 70°C) and mild centrifugation, the triglyceride content of the flies extract was determined using Sigma Triglyceride (INT 336-10 or -20) assay by measuring changes in the optical density according to the manufacturer's protocol. In case of the EP-lines EP(2)2449, HD-EP(3)31646, and HD-EP(3)30815, as a reference the protein content of the same extract was measured using BIO-RAD DC

Protein Assay according to the manufacturer's protocol. These assays were repeated at least three times.

The average triglyceride level of about 50 lines of the PX collection (referred to as 'PX-control') is shown as 1 (relative amount of triglyceride per fly) in the first column in FIGURE 1. The average triglyceride level of all flies of the EP collection (referred to as 'EP-control') is shown as 100% in the first columns in FIGURE 4, and 10. Standard deviations of the measurements are shown as thin bars.

PX9430.2 homozygous flies show constantly a higher triglyceride content than the controls (75%; column 2 in FIGURE 1, 'PX 9430.2'). Therefore, the loss of gene activity in the locus 11E6 on chromosome X where the EP-vector of PX9430.2 flies is homozygous viable integrated, is responsible for changes in the metabolism of the energy storage triglycerides, therefore representing a model for obese flies. The findings suggest the presence of similar functions of the homologous proteins in humans.

EP(2)2449 homozygous flies show constantly a lower triglyceride content than the controls (35%; column 2 in FIGURE 4, 'EP(2)2449'). Therefore, the loss of gene activity in the locus 29E3-4 on chromosome 2L where the EP-vector of EP(2)2449 flies is homozygous viable integrated, is responsible for changes in the metabolism of the energy storage triglycerides.

HD-EP(3)31646 males were crossed to elav-Gal4 virgins. The offspring carries a copy of the HD-EP(3)31646 vector and a copy of the elav-Gal4 vector, leading to ectopic expression of adjacent genomic DNA sequences 3' of the HD-EP(3)31646 integration locus, mainly in the neurons of these flies. The flies were analyzed in an assay measuring the triglyceride content of these flies. The result of the triglyceride content analysis is shown in FIGURE 7. The average triglyceride level of the fly collection in which the

HD-EP(3)31646 line was found, is crossed to elav-Gal4 flies, shown as 100% in FIGURE 7 (referred to as 'random EP/elav', column 1). HD-EP(3)31646 homozygous flies show constantly a higher triglyceride content than the controls (50%; column 2 in FIGURE 7). Therefore, the gain of gene activity in the locus 85D25 on chromosome 3R, where the EP-vector of HD-EP(3)31646 flies is homozygous viable integrated in the promoter of the CG9373 gene, is responsible for changes in the metabolism of the energy storage triglycerides.

HD-EP(3)30815 homozygous flies show constantly a lower triglyceride content than the controls (35%; column 2 in FIGURE 4, 'HD-EP30815'). Therefore, the loss of gene activity in the locus 94A1-2 on chromosome 3R where the EP-vector of HD-EP(3)30815 flies is homozygous viable integrated, is responsible for changes in the metabolism of the energy storage triglycerides.

Example 2: Identification of the genes

Genomic DNA sequences were isolated that are localized to the EP vector (herein PX9430.2) integration. Using those isolated genomic sequences public databases like Berkeley Drosophila Genome Project (GadFly) were screened thereby confirming the homozygous viable integration site of the PX9430.2 vector into the leader sequence of a Drosophila gene, identified as Jafrac1 (GadFly Accession Number CG1633). FIGURE 2 shows the molecular organization of this gene locus. The chromosomal localization site of the integration of the vector of PX9430.2 is at gene locus X, 11E6. In FIGURE 2, genomic DNA sequence is represented by the assembly as a scaled black line on the bottom, that includes the integration site of PX9430.2. The insertion site of the P-element in Drosophila PX9430.2 line is shown as vertical labeled line. Black boxes, linked by thin black lines predicted genes (as predicted by the Berkeley Drosophila Genome Project). Predicted exons are shown as black boxes, predicted introns are shown as

thin black lines. Transcribed DNA sequences (ESTs) are shown as grey bars. Therefore, expression of the cDNA encoding CG1633 could be affected by homozygous integration of vectors of line PX9430.2, leading to an increase of the energy storage triglycerides.

Genomic DNA sequences were isolated that are localized to the EP vector (herein EP(2)2449) integration. Using those isolated genomic sequences public databases like Berkeley Drosophila Genome Project (GadFly) were screened thereby confirming the homozygous viable integration site of the EP(2)2449 vector into an intron of a Drosophila gene in antisense orientation, identified as Hnf4 (GadFly Accession Number CG9310). FIGURE 5 shows the molecular organization of this gene locus. The chromosomal localization site of the integration of the vector of EP(2)2449 is at gene locus 2L, 29E3-4. In FIGURE 5, genomic DNA sequence is represented by the assembly as a thin black line in the middle that includes the integration sites of vector for line EP(2)2449. Ticks represent the legth of the genomic DNA in base pairs (1000 base pairs per tick). The insertion site of the P-element in Drosophila EP(2)2449 line is shown in the upper half as triangle and as vertical line. Grey Black boxes, linked by thin black lines represent the predicted genes (as predicted by the Berkeley Drosophila Genome Project). Predicted exons are shown as black boxes, predicted introns are shown as black lines. Transcribed DNA sequences (ESTs) are shown as grey bars. Therefore, expression of the cDNA encoding Hnf4 (Accession Number CG9310) could be affected by homozygous integration of vectors of line EP(2)2449, leading to a decrease of the energy storage triglycerides.

Genomic DNA sequences were isolated that are localized to the EP vector (herein HD-EP(3)31646) integration. Using those isolated genomic sequences public databases like Berkeley Drosophila Genome Project (GadFly) were screened thereby confirming the homozygous viable integration site of the HD-EP(3)31646 vector into the promoter region of a

Drosophila gene in sense orientation, identified as CG9373 (GadFly Accession Number). FIGURE 8 shows the molecular organization of this gene locus. The chromosomal localization site of the integration of the vector of HD-EP(3)31646 is at gene locus 3R, 85D25. In FIGURE 8, genomic DNA sequence is represented by the assembly as a dotted black line in the middle that includes the integration sites of vector for line HD-EP(3)31646. Numbers represent the coordinates of the genomic DNA (starting at position 5312505 on chromosome 3R, ending at position 5318755 on chromosome 3R). The insertion site of the P-element in Drosophila HD-EP(3)31646 line is shown in the lower "P-elements" line. Black boxes on the the lower "cDNA"-line, linked by light grey boxes, represent the predicted genes (as predicted by the Berkeley Drosophila Genome Project, GadFly and by Magpie). Predicted exons are shown as black boxes, predicted introns are shown as light grey boxes. The gene CG9373 is labeled. Transcribed DNA sequences (ESTs) are shown as grey bars in the lower "EST" line. Therefore, expression of the cDNA encoding CG9373 (GadFly Accession Number) could be affected by homozygous integration of vectors of line HD-EP(3)31646, leading to an increase of the energy storage triglycerides.

Genomic DNA sequences were isolated that are localized to the EP vector (herein HD-EP(3)30815) integration. Using those isolated genomic sequences public databases like Berkeley Drosophila Genome Project (GadFly) were screened thereby confirming the homozygous viable integration site of the HD-EP(3)30815 vector into the promoter of a Drosophila gene in antisense orientation, identified as how (GadFly Accession Number CG10293). FIGURE 11 shows the molecular organization of this gene locus. The chromosomal localization site of the integration of the vector of HD-EP(3)30815 is at gene locus 3R, 94A1-2. In FIGURE 11, genomic DNA sequence is represented by the assembly as a dotted black line in the middle that includes the integration sites of vector for line HD-EP(3)30815. Numbers represent the coordinates of the genomic

DNA (starting at position 17775577 on chromosome 3R, ending at position 17775577 on chromosome 3R). The insertion site of the P-element in Drosophila HD-EP(3)30815 line is shown in the lower "P-elements" line. Grey bars on the upper "cDNA"-line represent the predicted genes (as predicted by the Berkeley Drosophila Genome Project, GadFly and by Magpie). Predicted exons are shown as black boxes, predicted introns are shown as light grey boxes. The gene CG10293 is labeled. Transcribed DNA sequences (ESTs) are shown as grey bars in the upper "EST" lines. Therefore, expression of the cDNA encoding how (Accession Number CG10293) could be affected by homozygous integration of vectors of line HD-EP(3)30815, leading to a decrease of the energy storage triglycerides.

Example 3: Identification of human thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings homologous proteins

thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings homologous proteins and nucleic acid molecules coding therefore are obtainable from insect or vertebrate species, e.g. mammals or birds. Particularly preferred are nucleic acids comprising Jafrac1 (GadFly Accession Number CG1633), the human peroxiredoxin family, Hnf4 (GadFly Accession Number CG9310), the human hepatocyte nuclear factor 4 family, CG9373 (GadFly Accession Number), human myelin expression factor 2 and closely related proteins, how (GadFly Accession Number CG10293), and the human quaking isoforms.

As shown in FIGURE 3, gene product of GadFly Accession Number CG1633 is 83% homologous to human peroxiredoxin 2 (GenBank Accession Number XP_009063.2 for the protein, XM_009062 for the cDNA), 82% homologous to human peroxiredoxin 1 (GenBank Accession Number NP_002565.1 for the protein, NM_002574 for the cDNA), 83% homologous to thioredoxin peroxidase (identical to peroxiredoxin 4; GenBank Accession Number NP_006397.1 for the protein, NM_006406 for

the cDNA), and 77% to peroxiredoxin 3 (GenBank Accession Number NP_006784.1 for the protein, NM_006793 for the cDNA. CG1633 also shows 82% homology on protein level to mouse thioredoxin dependent peroxide reductase 2 (GenBank Accession Number NP_035164.1) and 83% homology on protein level to mouse peroxiredoxin 4 (GenBank Accession Number NP_048044.1).

As shown in FIGURE 6, gene product of GadFly Accession Number CG9310 is 72% homologous to human hepatocyte nuclear factor 4A (GenBank Accession Number CAA61133.1 for the protein, X87870 for the cDNA), 71% homologous to human hepatocye nuclear factor 4, splice form B (GenBank Accession Number CAA61134.1 for the protein, X87871 for the cDNA), 68% homologous to hepatocyte nuclear factor 4 alpha2 chain (GenBank Accession Number CAA89989.1 for the protein, Z49825 for the cDNA), and 68% to hepatocyte nuclear factor 4-alpha (GenBank Accession Number AAB48082.1 for the protein, U72969 for the cDNA. CG9310 also shows 77% homology on protein level to mouse hepatocye nuclear factor 4-alpha (GenBank Accession Number P49698).

As shown in FIGURE 9, gene product of GadFly Accession Number CG9373 is 44% homologous to human KIAA1341 protein (GenBank Accession Number BAA92579.1 for the protein, AB037762 for the cDNA), 43% homologous to human unnamed protein product (GenBank Accession Number BAB14421.1 for the protein, AK023133 for the cDNA), and 43% to myelin gene expression factor 2 (GenBank Accession Number NP_057216.1 for the protein, NM_016132 for the cDNA. CG9373 also shows 43% homology on protein level to mouse myelin gene expression factor (GenBank Accession Number AAL90778.1).

As shown in FIGURE 12, gene product of GadFly Accession Number CG10293 is 64% homologous to human QUAKING isoform 5 (GenBank Accession Number AAF63416.1 for the protein, AF142421 for the cDNA),

64% homologous to human protein similar to KH domain RNA binding protein QKI-5A (GenBank Accession Number XP_037438.2 for the protein, XM_037438 for the cDNA), 64% homologous to QUAKING isoform 6 (GenBank Accession Number AAF63414.1 for the protein, AF142419 for the cDNA), 64% homologous to unnamed protein product (GenBank Accession Number BAB55032.1 for the protein, AK027309 for the cDNA), 67% homologous to QUAKING isoform 2 (GenBank Accession Number AAF63413.1 for the protein, AF142418 for the cDNA), 67% homologous to QUAKING isoform 3 (GenBank Accession Number AAF63417.1 for the protein, AF142422 for the cDNA), 67% homologous to QUAKING isoform 4 (GenBank Accession Number AAF63415.1 for the protein, AF142420 for the cDNA), 67% homologous to QUAKING isoform 3 (GenBank Accession Number AAF63417.1 for the protein, AF142422 for the cDNA), 67% homologous to RNA binding protein HQK-6 (GenBank Accession Number BAB69497.1 for the protein, AB067799 for the cDNA), 67% homologous to RNA binding protein HQK-7B (GenBank Accession Number BAB69499.1 for the protein, AB067801 for the cDNA), 67% homologous to RNA binding protein HQK-7 (GenBank Accession Number BAB69498.1 for the protein, AB067800 for the cDNA), 67% homologous to QUAKING isoform 1 (GenBank Accession Number AAF63412.1 for the protein, AF142417 for the cDNA), and 64% to genes related to stomach cancer (GenBank Accession Number BD004960.1. CG10293 also shows 67% homology on protein level to mouse KH domain RNA binding protein QKI-7B (GenBank Accession Number AAC63042.1).

Claims

1. A pharmaceutical composition comprising a nucleic acid molecule of the thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings gene family or a polypeptide encoded thereby or a fragment or a variant of said nucleic acid molecule or said polypeptide or an antibody, an aptamer or another receptor recognizing a nucleic acid molecule of the thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings gene family or a polypeptide encoded thereby together with pharmaceutically acceptable carriers, diluents and/or adjuvants.

2. The composition of claim 1, wherein the nucleic acid molecule is a vertebrate or insect thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings nucleic acid, particulary a nucleic acid of the human peroxiredoxin family, a nucleic acid of the human hepatocyte nuclear factor 4 family, a nucleic acid of human myelin expression factor 2 gene and closely related gene or a nucleic acid of the human quaking isoform genes, and/or a nucleic molecule which is complementary thereto, or a fragment thereof or a variant thereof.

3. The composition of claim 1 or 2, wherein said nucleic acid molecule
   (a) hybridizes at 50°C in a solution containing 1 x SSC and 0.1% SDS to a nucleic acid molecule as defined in claim 2 and/or a nucleic acid molecule which is complementary thereto;
   (b) it is degenerate with respect to the nucleic acid molecule of (a),
   (c) encodes a polypeptide which is at least 85%, preferably at least 90%, more preferably at least 95%, more preferably at least 98% and up to 99,6% identical to a protein of the

human peroxiredoxin family, a protein of the human hepatocyte nuclear factor 4 family, human myelin expression factor 2 and closely related proteins, or the human quaking isoforms, as defined in claim 2;

(d) differs from the nucleic acid molecule of (a) to (c) by mutation and wherein said mutation causes an alteration, deletion, duplication or premature stop in the encoded polypeptide.

4. The composition of any one of claims 1-3, wherein the nucleic acid molecule is a DNA molecule, particularly a cDNA or a genomic DNA.

5. The composition of any one of claims 1-4, wherein said nucleic acid encodes a polypeptide contributing to regulating the energy homeostasis and/or the metabolism of triglycerides.

6. The composition of any one of claims 1-5, wherein said nucleic acid molecule is a recombinant nucleic acid molecule.

7. The composition of any one of claims 1-6, wherein the nucleic acid molecule is a vector, particularly an expression vector.

8. The composition of any one of claims 1-5, wherein the polypeptide is a recombinant polypeptide.

9. The composition of claim 8, wherein said recombinant polypeptide is a fusion polypeptide.

10. The composition of any one of claims 1-7, wherein said nucleic acid molecule is selected from hybridization probes, primers and anti-sense oligonucleotides.

11. The composition of any one of claims 1-10 which is a diagnostic composition.

12. The composition of any one of claims 1-10 which is a therapeutic composition.

13. The composition of any one of claims 1-12 for the manufacture of an agent for detecting and/or verifying, for the treatment, alleviation and/or prevention of an disorders, including metabolic diseases such as obesity and other body-weight regulation disorders as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apnea and others, in cells, cell masses, organs and/or subjects.

14. Use of a nucleic acid molecule of the thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings gene family or a polypeptide encoded thereby or a fragment or a variant of said nucleic acid molecule or said polypeptide or an antibody, an aptamer or another receptor recognizing a nucleic acid molecule of the thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings gene family or a polypeptide encoded thereby for controlling the function of a gene and/or a gene product which is influenced and/or modified by a thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings homologous polypeptide.

15. Use of the nucleic acid molecule of the thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings gene family or a polypeptide encoded thereby or a fragment or a variant of said nucleic acid molecule or said polypeptide or an antibody, an aptamer

or another receptor recognizing a nucleic acid molecule of the thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings gene family or a polypeptide encoded thereby for identifying substances capable of interacting with a thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings homologous polypeptide.

16. A non-human transgenic animal exhibiting a modified expression of a thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings homologous polypeptide.

17. The animal of claim 16, wherein the expression of the thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings homologous polypeptide is increased and/or reduced.

18. A recombinant host cell exhibiting a modified expression of a thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings homologous polypeptide.

19. The cell of claim 18 which is a human cell.

20. A method of identifying a (poly)peptide involved in the regulation of energy homeostasis and/or metabolism of triglycerides in a mammal comprising the steps of

(a) contacting a collection of (poly)peptides with a thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings homologous polypeptide or a fragment thereof under conditions that allow binding of said (poly)peptides;

(b) removing (poly)peptides which do not bind and

(c) identifying (poly)peptides that bind to said thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings homologous polypeptide.

21. A method of screening for an agent which modulates the interaction of a thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings homologous polypeptide with a binding target/agent, comprising the steps of

(a) incubating a mixture comprising

(aa) a thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings homologous polypeptide, or a fragment thereof;

(ab) a binding target/agent of said thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings homologous polypeptide or fragment thereof; and

(ac) a candidate agent

under conditions whereby said thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings polypeptide or fragment thereof specifically binds to said binding target/agent at a reference affinity;

(b) detecting the binding affinity of said thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings polypeptide or fragment thereof to said binding target to determine an (candidate) agent-biased affinity; and

(c) determining a difference between (candidate) agent-biased affinity and the reference affinity.

22. A method of producing a composition comprising the (poly)peptide identified by the method of claim 20 or the agent identified by the method of claim 21 with a pharmaceutically acceptable carrier, diluent and/or adjuvant.

23. The method of claim 22 wherein said composition is a pharmaceutical composition for preventing, alleviating or treating of diseases and disorders, including metabolic diseases such as obesity

and other body-weight regulation disorders as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apnea and other diseases and disorders.

24. Use of a (poly)peptide as identified by the method of claim 20 or of an agent as identified by the method of claim 21 for the preparation of a pharmaceutical composition for the treatment, alleviation and/or prevention of of diseases and disorders, including metabolic diseases such as obesity and other body-weight regulation disorders as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apnea and other diseases and disorders.

25. Use of a nucleic acid molecule of the thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings family or of a fragment thereof for the preparation of a non-human animal which over- or under-expresses the thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings gene product.

26. Kit comprising at least one of

(a) a thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings nucleic acid molecule or a fragment thereof;

(b) a vector comprising the nucleic acid of (a);

(c) a host cell comprising the nucleic acid of (a) or the vector of (b);

(d) a polypeptide encoded by the nucleic acid of (a);

(e) a fusion polypeptide encoded by the nucleic acid of (a);

(f) an antibody, an aptamer or another receptor against the nucleic acid of (a) or the polypeptide of (d) or (e) and

(g) an anti-sense oligonucleotide of the nucleic acid of (a).

Abstract

The present invention discloses thioredoxin peroxidase 1, hepatocyte nuclear factor 4, CG9373, or held out wings homologous proteins regulating the energy homeostasis and the metabolism of triglycerides, and polynucleotides, which identify and encode the proteins disclosed in this invention. The invention also relates to the use of these sequences in the diagnosis, study, prevention, and treatment of diseases and disorders, for example, but not limited to, metabolic diseases such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancers of the reproductive organs, and sleep apnea.

cl 07.05.2002

FIGURE 1. Triglyceride content of a Drosophila *Jafrac1* (GadFly Accession Number CG1633) mutant

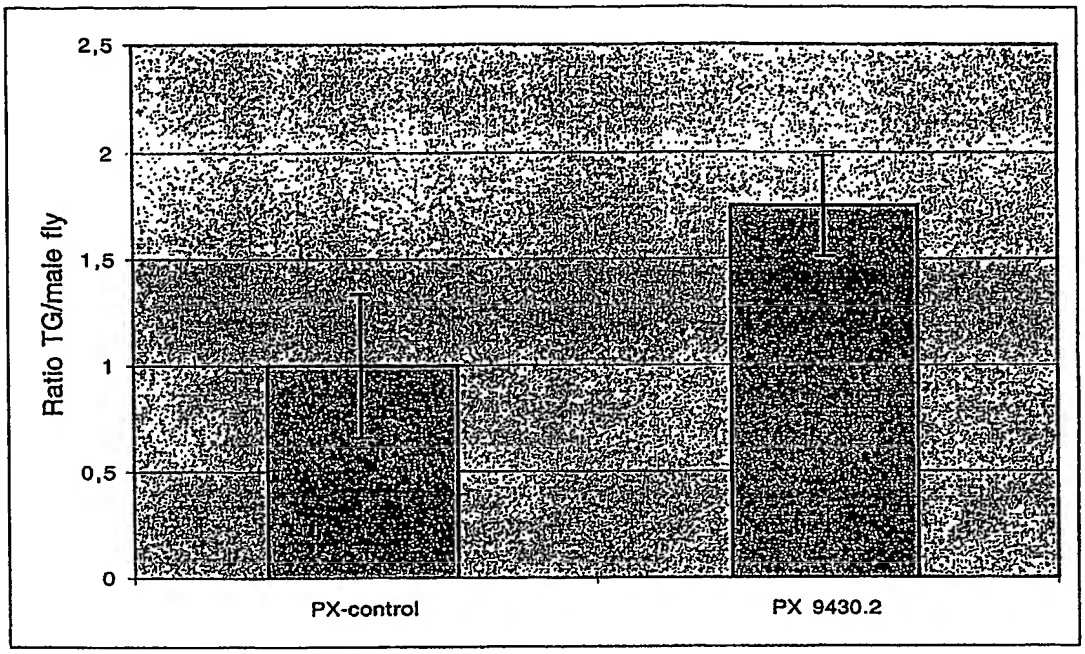

FIGURE 2. Molecular organisation of the *Jafrac1* gene (GadFly Accession Number CG1633)

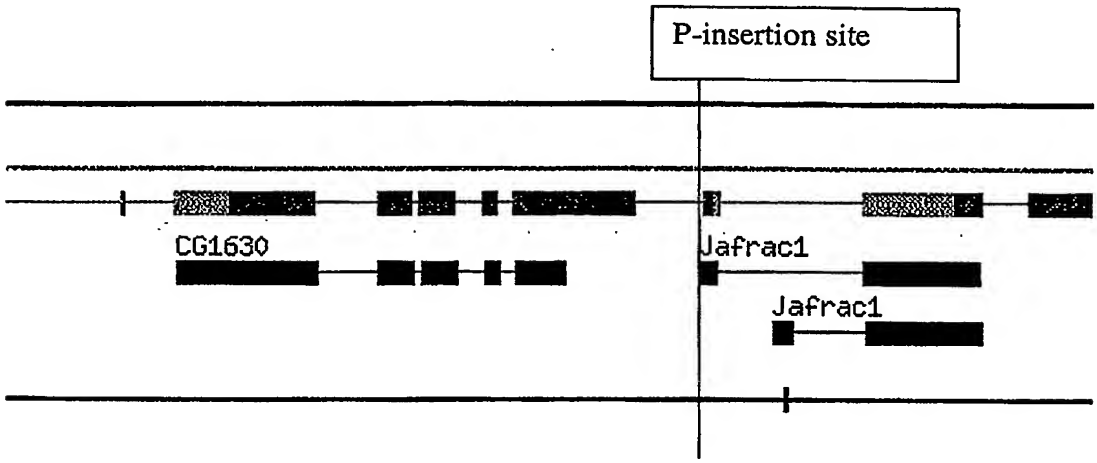

FIGURE 3. BLASTP results for CG1633 (GadFly Accession Number)

Homology to human protein XP_009063.2 (GenBank Accession Number)

```
ref|XP_009063.2| (XM_009063) peroxiredoxin 2 [Homo sapiens]
Length = 198

Score =  283 bits (723), Expect = 9e-76
Identities = 134/188 (71%), Positives = 157/188 (83%)

Query: 3   QLQKPAPAFAGTAVVNGVFKDIKLSDYKGKYLVLFFYPLDFTFVCPTEIIAFSESAAEFR 62
           ++ KPAP F  TAVV+G FK++KLSDYKGKY+VLFFYPLDFTFVCPTEIIAFS  A +FR
Sbjct: 7   RIGKPAPDFKATAVVDGAFKEVKLSDYKGKYVVLFFYPLDFTFVCPTEIIAFSNRAEDFR 66

Query: 63  KINCEVIGCSTDSQFTHLAWINTPRKQGGLGSMDIPLLADKSMKVARDYGVLDEETGIPF 122
           K+ CEV+G S DSQFTHLAWINTPRK+GGLG ++IPLLAD + +++ DYGVL  + GI +
Sbjct: 67  KLGCEVLGVSVDSQFTHLAWINTPRKEGGLGPLNIPLLADVTRRLSEDYGVLKTDEGIAY 126

Query: 123 RGLFIIDDKQNLRQITVNDLPVGRSVEETLRLVQAFQYTDKYGEVCPANWKPGQKTMVAD 182
           RGLFIID K  LRQITVNDLPVGRSV+E LRLVQAFQYTD++GEVCPA WKPG  T+  +
Sbjct: 127 RGLFIIDGKGVLRQITVNDLPVGRSVDEALRLVQAFQYTDEHGEVCPAGWKPGSDTIKPN 186

Query: 183 PTKSKEYF 190
              SKEYF
Sbjct: 187 VDDSKEYF 194
```

Homology to human protein NP_002565.1 (GenBank Accession Number)

```
ref|NP_002565.1| (NM_002574) peroxiredoxin 1; Proliferation-associated gene A;
proliferation-associated gene A (natural killer-enhancing factor A) [Homo sapiens]
ref|XP_001393.2| (XM_001393) peroxiredoxin 1 [Homo sapiens]
Length = 199

Score =  281 bits (718), Expect = 3e-75
Identities = 135/185 (72%), Positives = 154/185 (82%), Gaps = 1/185 (0%)

Query: 7   PAPAFAGTAVV-NGVFKDIKLSDYKGKYLVLFFYPLDFTFVCPTEIIAFSESAAEFRKIN 65
           PAP F  TAV+ +G FKDI LSDYKGKY+V FFYPLDFTFVCPTEIIAFS+ A EF+K+N
Sbjct: 11  PAPNFKATAVMPDGQFKDISLSDYKGKYVVFFFYPLDFTFVCPTEIIAFSDRAEEFKKLN 70

Query: 66  CEVIGCSTDSQFTHLAWINTPRKQGGLGSMDIPLLADKSMKVARDYGVLDEETGIPFRGL 125
           C+VIG S DS F HLAW+NTP+KQGGLG M+IPL++D     +A+DYGVL  + GI FRGL
Sbjct: 71  CQVIGASVDSHFCHLAWVNTPKKQGGLGPMNIPLVSDPKRTIAQDYGVLKADEGISFRGL 130

Query: 126 FIIDDKQNLRQITVNDLPVGRSVEETLRLVQAFQYTDKYGEVCPANWKPGQKTMVADPTK 185
           FIIDDK  LRQITVNDLPVGRSV+ETLRLVQAFQ+TDK+GEVCPA WKPG  T+  D  K
Sbjct: 131 FIIDDKGILRQITVNDLPVGRSVDETLRLVQAFQFTDKHGEVCPAGWKPGSDTIKPDVQK 190

Query: 186 SKEYF 190
           SKEYF
Sbjct: 191 SKEYF 195
```

Homology to human protein NP_006397.1 (GenBank Accession Number)

```
>ref|NP_006397.1| (NM_006406) thioredoxin peroxidase; thioredoxin peroxidase
(antioxidant enzyme) [Homo sapiens]
Length = 271

Score =  280 bits (715), Expect = 7e-75
Identities = 129/189 (68%), Positives = 158/189 (83%)

Query: 3   QLQKPAPAFAGTAVVNGVFKDIKLSDYKGKYLVLFFYPLDFTFVCPTEIIAFSESAAEFR 62
           ++ KPAP + GTAV++G FK++KL+DY+GKYLV FFYPLDFTFVCPTEIIAF +   EFR
Sbjct: 80  KISKPAPYWEGTAVIDGEFKELKLTDYRGKYLVFFFYPLDFTFVCPTEIIAFGDRLEEFR 139

Query: 63  KINCEVIGCSTDSQFTHLAWINTPRKQGGLGSMDIPLLADKSMKVARDYGVLDEETGIPF 122
            IN EV+ CS DSQFTHLAWINTPR+QGGLG + IPLL+D + ++++DYGV  E++G
Sbjct: 140 SINTEVVACSVDSQFTHLAWINTPRRQGGLGPIRIPLLSDLTHQISKDYGVYLEDSGHTL 199

Query: 123 RGLFIIDDKQNLRQITVNDLPVGRSVEETLRLVQAFQYTDKYGEVCPANWKPGQKTMVAD 182
           RGLFIIDDK  LRQIT+NDLPVGRSV+ETLRLVQAFQYTDK+GEVCPA WKPG +T++ D
Sbjct: 200 RGLFIIDDKGILRQITLNDLPVGRSVDETLRLVQAFQYTDKHGEVCPAGWKPGSETIIPD 259

Query: 183 PTKSKEYFE 191
           P    +YF+
Sbjct: 260 PAGKLKYFD 268
```

Homology to human protein NP_006784.1 (GenBank Accession Number)

```
ref|NP_006784.1| (NM_006793) peroxiredoxin 3; antioxidant protein 1;
thioredoxin-dependent peroxide reductase precursor [Homo sapiens]
ref|XP_055573.1| (XM_055573) peroxiredoxin 3 [Homo sapiens]
Length = 256

Score =  263 bits (671), Expect = 9e-70
Identities = 123/190 (64%), Positives = 149/190 (77%)

Query: 2   PQLQKPAPAFAGTAVVNGVFKDIKLSDYKGKYLVLFFYPLDFTFVCPTEIIAFSESAAEF 61
           P + + AP F GTAVVNG FKD+ L D+KGKYLVLFFYPLDFTFVCPTEI+AFS+ A EF
Sbjct: 63  PAVTQHAPYFKGTAVVNGEFKDLSLDDFKGKYLVLFFYPLDFTFVCPTEIVAFSDKANEF 122

Query: 62  RKINCEVIGCSTDSQFTHLAWINTPRKQGGLGSMDIPLLADKSMKVARDYGVLDEETGIP 121
             +NCEV+  S DS F+HLAWINTPRK GGLG M+I LL+D + +++RDYGVL E +G+
Sbjct: 123 HDVNCEVVAVSVDSHFSHLAWINTPRKNGGLGHMNIALLSDLTKQISRDYGVLLEGSGLA 182

Query: 122 FRGLFIIDDKQNLRQITVNDLPVGRSVEETLRLVQAFQYTDKYGEVCPANWKPGQKTMVA 181
            RGLFIID     ++ ++VNDLPVGRSVEETLRLV+AFQY + +GEVCPANW P    T+
Sbjct: 183 LRGLFIIDPNGVIKHLSVNDLPVGRSVEETLRLVKAFQYVETHGEVCPANWTPDSPTIKP 242

Query: 182 DPTKSKEYFE 191
            P  SKEYF+
Sbjct: 243 SPAASKEYFQ 252
```

FIGURE 4. Triglyceride content of a Drosophila *Hnf4* (GadFly Accession Number CG9310) mutant

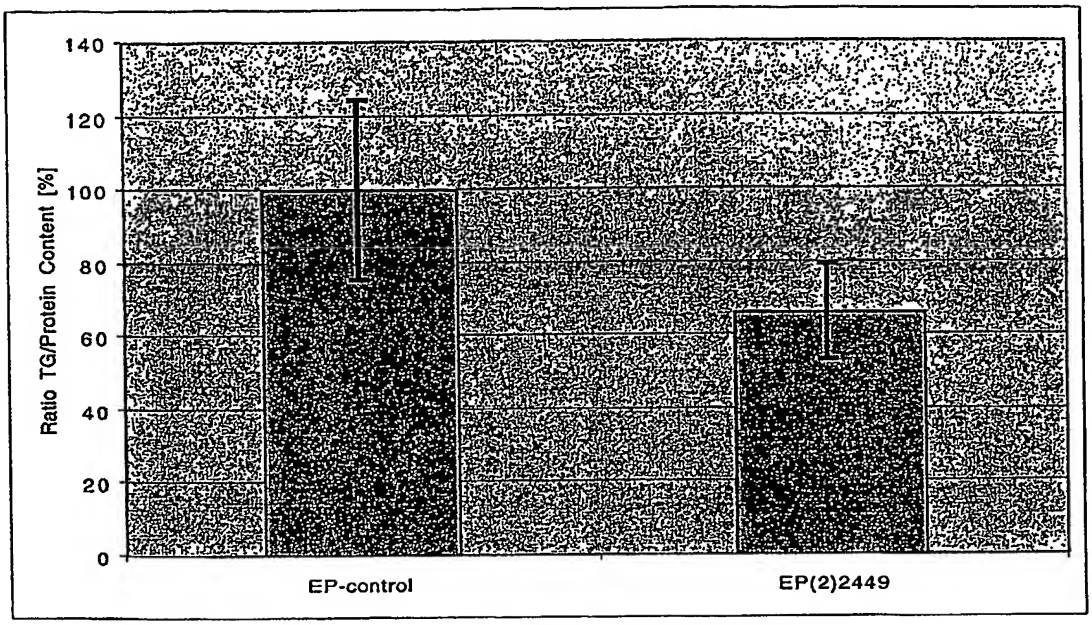

FIGURE 5. Molecular organisation of the *Hnf4* gene (GadFly Accession Number CG9310)

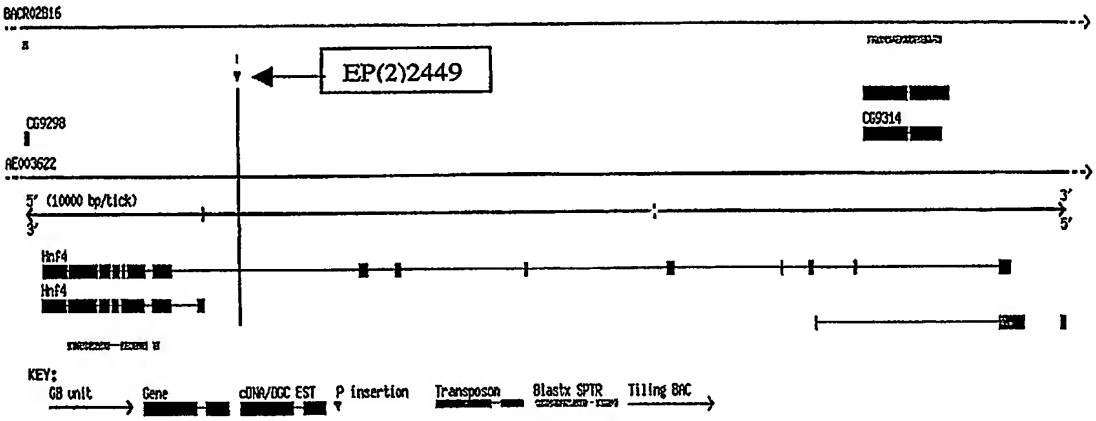

FIGURE 6. BLASTP results for CG9310 (GadFly Accession Number)

Homology to human protein JC4936 (GenBank Accession Number)

```
pir||JC4936 hepatocyte nuclear factor 4A - human
emb|CAA61133.1| (X87870) Hepatocyte nuclear factor 4A [Homo sapiens]
Length = 455

Score =  455 bits (1171), Expect = e-127
Identities = 241/408 (59%), Positives = 297/408 (72%), Gaps = 25/408 (6%)

Query: 103 VCAICGDRATGKHYGASSCDGCKGFFRRSVRKNHQYTCRFARNCVVDKDKRNQCRYCRLR 162
           +CAICGDRATGKHYGASSCDGCKGFFRRSVRKNH Y+CRF+R CVVDKDKRNQCRYCRL+
Sbjct: 50  LCAICGDRATGKHYGASSCDGCKGFFRRSVRKNHMYSCRFSRQCVVDKDKRNQCRYCRLK 109

Query: 163 KCFKAGMKKEAVQNERDRISCRRTSNDDPDPGNGLSVISLVKAENESRQSKAGAAMEPNI 222
           KCF+AGMKKEAVQNERDRIS RR+S +D    +  S+ +L++AE  SRQ  +  +    I
Sbjct: 110 KCFRAGMKKEAVQNERDRISTRRSSYED---SSLPSINALLQAEVLSRQITSPVS---GI 163

Query: 223 NEDLSNKQFASINDVCESMKQQLLTLVEWAKQIPAFNELQLDDQVALLRAHAGEHLLLGL 282
           N D+  K+ ASI DVCESMK+QLL LVEWAK IPAF EL LDDQVALLRAHAGEHLLLG
Sbjct: 164 NGDIRAKKIASIADVCESMKEQLLVLVEWAKYIPAFCELPLDDQVALLRAHAGEHLLLGA 223

Query: 283 SRRSMHLKDVLLLSNNCVITRHCPDPLVSPNLDISRIGARIIDELVTVMKDVGIDDTEFA 342
           ++RSM  KDVLLL N+ ++ RHCP+       ++SR+  RI+DELV   +++ IDD E+A
Sbjct: 224 TKRSMVFKDVLLLGNDYIVPRHCPE-----LAEMSRVSIRILDELVLPFQELQIDDNEYA 278

Query: 343 CIKALVFFDPNAKGLNEPHRIKSLRHQILNNLEDYISDRQYESRGRFGEILLILPVLQSI 402
            +KA++FFDP+AKGL++P +IK LR Q+  +LEDYI+DRQY+SRGRFGE+LL+LP LQSI
Sbjct: 279 YLKAIIFFDPDAKGLSDPGKIKRLRSQVQVSLEDYINDRQYDSRGRFGELLLLLPTLQSI 338

Query: 403 TWQMIEQIQFAKIFGVAHIDSLLQEMLLGGELADNPLPLSP--PNQSNDYQS-------- 452
           TWQMIEQIQF K+FG+A ID+LLQEMLLGG  +D P    P  P+   ++
Sbjct: 339 TWQMIEQIQFIKLFGMAKIDNLLQEMLLGGSPSDAPHAHHPLHPHLMQEHMGTNVIVANT 398

Query: 453 -PTHTGNMEGGNQVNSSLDSLATSGGPGSHSLDLEVQHIQALIEANSA 499
            PTH  N   G          +  GG GS S  L    +  +++  SA
Sbjct: 399 MPTHLSN---GQMSTPETPQPSPPGGSGSESYKLLPGAVATIVKPLSA 443
```

Homology to human protein JC4937 (GenBank Accession Number)

```
pir||JC4937 hepatocyte nuclear factor 4, splice form B - human
emb|CAA61134.1| (X87871) Hepatocyte nuclear factor 4B [Homo sapiens]
Length = 465

Score =  453 bits (1165), Expect = e-126
Identities = 241/415 (58%), Positives = 298/415 (71%), Gaps = 29/415 (6%)

Query: 103 VCAICGDRATGKHYGASSCDGCKGFFRRSVRKNHQYTCRFARNCVVDKDKRNQCRYCRLR 162
           +CAICGDRATGKHYGASSCDGCKGFFRRSVRKNH Y+CRF+R CVVDKDKRNQCRYCRL+
Sbjct: 50  LCAICGDRATGKHYGASSCDGCKGFFRRSVRKNHMYSCRFSRQCVVDKDKRNQCRYCRLK 109

Query: 163 KCFKAGMKKEAVQNERDRISCRRTSNDDPDPGNGLSVISLVKAENESRQSKAGAAMEPNI 222
           KCF+AGMKKEAVQNERDRIS RR+S +D    +  S+ +L++AE  SRQ  +  +    I
Sbjct: 110 KCFRAGMKKEAVQNERDRISTRRSSYED---SSLPSINALLQAEVLSRQITSPVS---GI 163

Query: 223 NEDLSNKQFASINDVCESMKQQLLTLVEWAKQIPAFNELQLDDQVALLRAHAGEHLLLGL 282
           N D+  K+ ASI DVCESMK+QLL LVEWAK IPAF EL LDDQVALLRAHAGEHLLLG
Sbjct: 164 NGDIRAKKIASIADVCESMKEQLLVLVEWAKYIPAFCELPLDDQVALLRAHAGEHLLLGA 223

Query: 283 SRRSMHLKDVLLLSNNCVITRHCPDPLVSPNLDISRIGARIIDELVTVMKDVGIDDTEFA 342
```

```
            ++RSM  KDVLLL N+ ++ RHCP+        ++SR+  RI+DELV   +++ IDD E+A
Sbjct: 224 TKRSMVFKDVLLLGNDYIVPRHCPE-----LAEMSRVSIRILDELVLPFQELQIDDNEYA 278

Query: 343 CIKALVFFDPNAKGLNEPHRIKSLRHQILNNLEDYISDRQYESRGRFGEILLILPVLQSI 402
            +KA++FFDP+AKGL++P +IK LR Q+  +LEDYI+DRQY+SRGRFGE+LL+LP LQSI
Sbjct: 279 YLKAIIFFDPDAKGLSDPGKIKRLRSQVQVSLEDYINDRQYDSRGRFGELLLLLPTLQSI 338

Query: 403 TWQMIEQIQFAKIFGVAHIDSLLQEMLLGGELADNPLPLSP--PNQSNDYQS-------- 452
            TWQMIEQIQF K+FG+A ID+LLQEMLLGG  +D P    P  P+   ++
Sbjct: 339 TWQMIEQIQFIKLFGMAKIDNLLQEMLLGGSPSDAPHAHHPLHPHLMQEHMGTNVIVANT 398

Query: 453 -PTHTGNME-------GGNQVNSSLDSLATSGGPGSHSLDLEVQHIQALIEANSA 499
             PTH  N +        G          +  GG GS S  L    +  +++  SA
Sbjct: 399 MPTHLSNGQMCEWPRPRGQAATPETPQPSPPGGSGSESYKLLPGAVATIVKPLSA 453
```

Homology to human protein JC6096 (GenBank Accession Number)

```
pir||JC6096 hepatocyte nuclear factor 4 alpha2 chain - human
emb|CAA89989.1| (Z49825) hepatocyte nuclear factor 4 alpha (HNF4alpha4) [Homo sapiens]
Length = 504

Score =  451 bits (1161), Expect = e-125
Identities = 244/439 (55%), Positives = 302/439 (68%), Gaps = 29/439 (6%)

Query: 79  SGSGSGTNSSQQQLQQQQQQQSPTVCAICGDRATGKHYGASSCDGCKGFFRRSVRKNHQY 138
            S  G  + S    L           +CAICGDRATGKHYGASSCDGCKGFFRRSVRKNH Y
Sbjct: 65  SPQGDTSPSEGTNLNAPNSLGVSALCAICGDRATGKHYGASSCDGCKGFFRRSVRKNHMY 124

Query: 139 TCRFARNCVVDKDKRNQCRYCRLRKCFKAGMKKEAVQNERDRISCRRTSNDDPDPGNGLS 198
            +CRF+R CVVDKDKRNQCRYCRL+KCF+AGMKKEAVQNERDRIS RR+S +D    +  S
Sbjct: 125 SCRFSRQCVVDKDKRNQCRYCRLKKCFRAGMKKEAVQNERDRISTRRSSYED---SSLPS 181

Query: 199 VISLVKAENESRQSKAGAAMEPNINEDLSNKQFASINDVCESMKQQLLTLVEWAKQIPAF 258
            + +L++AE  SRQ  + ¥  IN D+ K+ ASI DVCESMK+QLL LVEWAK.IPAF
Sbjct: 182 INALLQAEVLSRQITSPVS---GINGDIRAKKIASIADVCESMKEQLLVLVEWAKYIPAF 238

Query: 259 NELQLDDQVALLRAHAGEHLLLGLSRRSMHLKDVLLLSNNCVITRHCPDPLVSPNLDISR 318
             EL LDDQVALLRAHAGEHLLLG ++RSM  KDVLLL N+ ++ RHCP+        ++SR
Sbjct: 239 CELPLDDQVALLRAHAGEHLLLGATKRSMVFKDVLLLGNDYIVPRHCPE-----LAEMSR 293

Query: 319 IGARIIDELVTVMKDVGIDDTEFACIKALVFFDPNAKGLNEPHRIKSLRHQILNNLEDYI 378
            +  RI+DELV   +++ IDD E+A +KA++FFDP+AKGL++P +IK LR Q+  +LEDYI
Sbjct: 294 VSIRILDELVLPFQELQIDDNEYAYLKAIIFFDPDAKGLSDPGKIKRLRSQVQVSLEDYI 353

Query: 379 SDRQYESRGRFGEILLILPVLQSITWQMIEQIQFAKIFGVAHIDSLLQEMLLGGELADNP 438
            +DRQY+SRGRFGE+LL+LP LQSITWQMIEQIQF K+FG+A ID+LLQEMLLGG  +D P
Sbjct: 354 NDRQYDSRGRFGELLLLLPTLQSITWQMIEQIQFIKLFGMAKIDNLLQEMLLGGSPSDAP 413

Query: 439 LPLSP--PNQSNDYQS---------PTHTGNME-------GGNQVNSSLDSLATSGGPGS 480
                P  P+   ++          PTH  N +        G          +  GG GS
Sbjct: 414 HAHHPLHPHLMQEHMGTNVIVANTMPTHLSNGQMCEWPRPRGQAATPETPQPSPPGGSGS 473

Query: 481 HSLDLEVQHIQALIEANSA 499
                L    +  +++  SA
Sbjct: 474 EPYKLLPGAVATIVKPLSA 492
```

Homology to human protein AAB48082.1 (GenBank Accession Number)

```
gb|AAB48082.1| (U72969) hepatocyte nuclear factor 4-alpha [Homo sapiens]
Length = 516

Score =  451 bits (1161), Expect = e-125
Identities = 244/439 (55%), Positives = 302/439 (68%), Gaps = 29/439 (6%)

Query: 79  SGSGSGTNSSQQQLQQQQQQQSPTVCAICGDRATGKHYGASSCDGCKGFFRRSVRKNHQY 138
           S  G  + S    L           +CAICGDRATGKHYGASSCDGCKGFFRRSVRKNH Y
Sbjct: 77  SPQGDTSPSEGTNLNAPNSLGVSALCAICGDRATGKHYGASSCDGCKGFFRRSVRKNHMY 136

Query: 139 TCRFARNCVVDKDKRNQCRYCRLRKCFKAGMKKEAVQNERDRISCRRTSNDDPDPGNGLS 198
           +CRF+R CVVDKDKRNQCRYCRL+KCF+AGMKKEAVQNERDRIS RR+S +D    +  S
Sbjct: 137 SCRFSRQCVVDKDKRNQCRYCRLKKCFRAGMKKEAVQNERDRISTRRSSYED---SSLPS 193

Query: 199 VISLVKAENESRQSKAGAAMEPNINEDLSNKQFASINDVCESMKQQLLTLVEWAKQIPAF 258
           + +L++AE  SRQ  +  +    IN D+  K+ ASI DVCESMK+QLL LVEWAK IPAF
Sbjct: 194 INALLQAEVLSRQITSPVS---GINGDIRAKKIASIADVCESMKEQLLVLVEWAKYIPAF 250

Query: 259 NELQLDDQVALLRAHAGEHLLLGLSRRSMHLKDVLLLSNNCVITRHCPDPLVSPNLDISR 318
            EL LDDQVALLRAHAGEHLLLG ++RSM  KDVLLL N+ ++ RHCP+       ++SR
Sbjct: 251 CELPLDDQVALLRAHAGEHLLLGATKRSMVFKDVLLLGNDYIVPRHCPE-----LAEMSR 305

Query: 319 IGARIIDELVTVMKDVGIDDTEFACIKALVFFDPNAKGLNEPHRIKSLRHQILNNLEDYI 378
           +  RI+DELV   +++ IDD E+A +KA++FFDP+AKGL++P +IK LR Q+  +LEDYI
Sbjct: 306 VSIRILDELVLPFQELQIDDNEYAYLKAIIFFDPDAKGLSDPGKIKRLRSQVQVSLEDYI 365

Query: 379 SDRQYESRGRFGEILLILPVLQSITWQMIEQIQFAKIFGVAHIDSLLQEMLLGGELADNP 438
           +DRQY+SRGRFGE+LL+LP LQSITWQMIEQIQF K+FG+A ID+LLQEMLLGG  +D P
Sbjct: 366 NDRQYDSRGRFGELLLLLPTLQSITWQMIEQIQFIKLFGMAKIDNLLQEMLLGGSPSDAP 425

Query: 439 LPLSP--PNQSNDYQS---------PTHTGNME-------GGNQVNSSLDSLATSGGPGS 480
               P  P+   ++            PTH  N +         G           +  GG GS
Sbjct: 426 HAHHPLHPHLMQEHMGTNVIVANTMPTHLSNGQMCEWPRPRGQAATPETPQPSPPGGSGS 485

Query: 481 HSLDLEVQHIQALIEANSA 499
               L    +  +++  SA
Sbjct: 486 EPYKLLPGAVATIVKPLSA 504
```

FIGURE 7. Triglyceride content of a Drosophila CG9373 (GadFly Accession Number) mutant

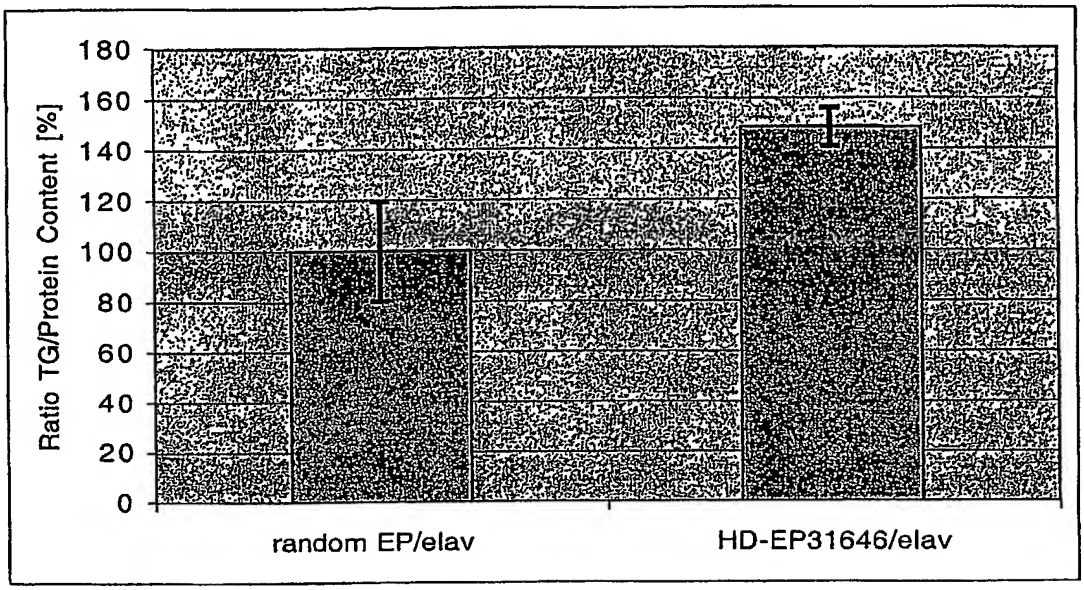

FIGURE 8. Molecular organisation of the CG9373 gene (GadFly Accession Number)

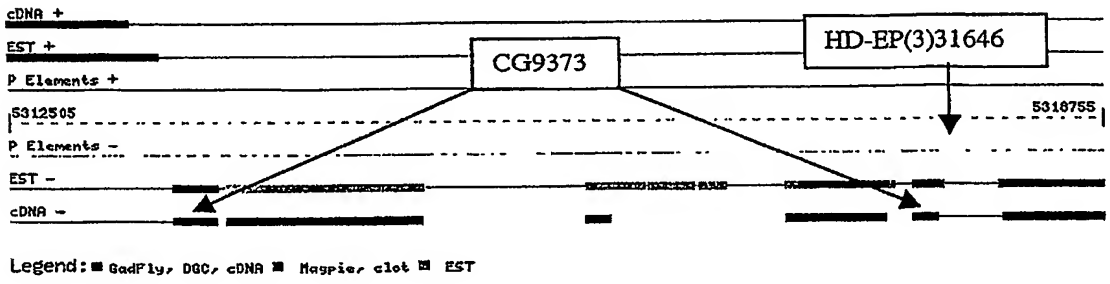

FIGURE 9. BLASTP results for CG9373 (GadFly Accession Number)

Homology to human protein BAA92579.1 (GenBank Accession Number)

```
dbj|BAA92579.1| (AB037762) KIAA1341 protein [Homo sapiens], Length = 620

Score =  249 bits (635), Expect = 1e-64
Identities = 207/660 (31%), Positives = 295/660 (44%), Gaps = 148/660 (22%)

Query: 1   MSMDASNSVESREKERDRRGRGAR-GSRFTDADGNGN-GAGSQGGGVAARDRSRERRNCR 58
           + M+   S +  + +   +  G++  +RF     + N G G + G           RN R
Sbjct: 72  VKMENDESAKEEKSDLKEKSTGSKKANRFHPYSKDKNSGTGEKKG---------PNRN-R 121

Query: 59  VYISNIPYDYRWQDLKDLFRRIVGSIEYVQLFFDESGKARGCGIVEFKDPENVQKALEKM 118
           V+ISNIPYD +WQ +KDL R  VG + YV+LF D  GK+RGCG+VEFKD E V+KALE M
Sbjct: 122 VFISNIPYDMKWQAIKDLMREKVGEVTYVELFKDAEGKSRGCGVVEFKDEEFVKKALETM 181

Query: 119 NRYEVNGRELVVKEDHGEQRDQYGRIVRDGGGGGGGGGGVQGGNGGNNGGGGGGGRDHMD 178
           N+Y+++GR L +KED   +  +   + R GG   GG                   H+
Sbjct: 182 NKYDLSGRPLNIKEDPDGENARRA-LQRTGGSFPGG---------------------HVP 219

Query: 179 DRDRGFSRRDDDRLSGRNNFNMMSNDYNNSSNYNLYGLSASFLESLGISGPLHNKVFVAN 238
           D  G        L   NN N+     +N                   +G L + +FVAN
Sbjct: 220 DMGSGLMNLPPSIL---NNPNIPPEVISNLQ-----------------AGRLGSTIFVAN 259

Query: 239 LDYKVDNKKLKQVFKLAGKVQSVDLSLDKEGNSRGFAVIEYDHPVEAVQAISMLDRQMLF 298
           LD+KV  KKLK+VF +AG V+  D+  DK+G SRG   + ++  +EAVQAISM + Q LF
Sbjct: 260 LDFKVGWKKLKEVFSIAGTVKRADIKEDKDGKSRGMGTVTFEQAIEAVQAISMFNGQFLF 319

Query: 299 DRRMTVRLD--RIPDK------NEGIKLPEGLGGVGIGLGPNGEPLRDVAHNLPNGGQSQ 350
           DR M V++D   +P +       +  +LP GLGG+G+GLGP G+P+     N+
Sbjct: 320 DRPMHVKMDDKSVPHEEYRSHDGKTPQLPRGLGGIGMGLGPGGQPISASQLNI------- 372

Query: 351 GQLLGNAQQGSQLGSVGSQPNSSAVSNATTNLLNNLTGVMFGNHAAVQPSPVAPVQKPSL 410
           G ++GN   G                    + G  FG                   +
Sbjct: 373 GGVMGNLGPGGM---------------------GMDGPGFGG-------------MNRI 397

Query: 411 GNNTGSGGLNLNNLNPSILAAVVGNLGNQG--GNLSNPLLSSSL------SNLGLNLGNS 462
           G   G GGL   N          +G  G  G  G L    ++SS+       ++G+N G
Sbjct: 398 GGGIGFGGLEAMN--------SMGGFGGVGRMGELYRGAMTSSMERDFGRGDIGINRGFG 449

Query: 463 GNDDNLPPSNVGLSNNYSSGGTGGGNSYSSGNNYSGGGGSSN-------LGYNAYSSS-G 514
            +   L  + +G      +G  G  N    G+  SGG GS N       +G +  SSS
Sbjct: 450 DSFGRLGSAMIG----GFAGRIGSSNMGPVGSGISGGMGSMNSVTGGMGMGLDRMSSSFD 505

Query: 515 GMGGGNGGVGVDGNDYNTGNPLDVYGGGSNVGNSNVGSANAVGASRKSDTIIIKNVPITC 574
            MG G G +     D + G      G G       +GS       K + I ++N+P
Sbjct: 506 RMGPGIGAILERSIDMDRGFLSGPMGSGM---RERIGS--------KGNQIFVRNLPFDL 554

Query: 575 TWQTLRDKFREIGDVKFAEI-----RGNDVGVVRFFKERDAELAIALMDGSRLDGRNIKV 629
           TWQ L++KF + G V FAEI     +    G VRF     AE A  +M+G ++ GR I V
Sbjct: 555 TWQKLKEKFSQCGHVMFAEIKMENGKSKGCGTVRFDSPESAEKACRIMNGIKISGREIDV 614

Score = 68.6 bits (166), Expect = 2e-10
Identities = 41/114 (35%), Positives = 67/114 (57%), Gaps = 6/114 (5%)

Query: 20  GRGARGSRFTDADGNGNGAGSQGGGVAARDRSRERRNCRVYISNIPYDYRWQDLKDLFRR 79
           G GA   R  D D  G  +G  G G+  R+R   + N ++++ N+P+D  WQ LK+ F +
Sbjct: 510 GIGAILERSIDMD-RGFLSGPMGSGM--RERIGSKGN-QIFVRNLPFDLTWQKLKEKFSQ 565

Query: 80  IVGSIEYVQLFFDESGKARGCGIVEFKDPENVQKALEKMNRYEVNGRELVVKED 133
             G + + ++   E+GK++GCG V F   PE+ +KA    MN  ++++GRE+ V+ D
Sbjct: 566 C-GHVMFAEIKM-ENGKSKGCGTVRFDSPESAEKACRIMNGIKISGREIDVRLD 617
```

```
Score = 56.2 bits (134), Expect = 1e-06
Identities = 46/180 (25%), Positives = 76/180 (41%), Gaps = 21/180 (11%)

Query: 139 DQYGRIVRDGGGGGGG----------GGGVQGGNGGNNGGGGGGGRDHMDDRDRGFSRRD 188
           D +GR+     GG  G          G G+ GG G  N   GG G   +D     F R
Sbjct: 450 DSFGRLGSAMIGGFAGRIGSSNMGPVGSGISGGMGSMNSVTGGMGMG-LDRMSSSFDRM- 507

Query: 189 DDRLSGRNNFNMMSNDYNNSSNYNLYGLSASFLESLGISGPLHNKVFVANLDYKVDNKKL 248
                G     ++    +    + +   E +G  G   N++FV NL + +  +KL
Sbjct: 508 -----GPGIGAILERSIDMDRGFLSGPMGSGMRERIGSKG---NQIFVRNLPFDLTWQKL 559

Query: 249 KQVFKLAGKVQSVDLSLDKEGNSRGFAVIEYDHPVEAVQAISMLDRQMLFDRRMTVRLDR 308
           K+ F   G V   ++ ++  G S+G   + +D P  A +A   +++   +  R + VRLDR
Sbjct: 560 KEKFSQCGHVMFAEIKMEN-GKSKGCGTVRFDSPESAEKACRIMNGIKISGREIDVRLDR 618
```

Homology to human protein BAB14421.1 (GenBank Accession Number)

```
>dbj|BAB14421.1| (AK023133) unnamed protein product [Homo sapiens], Length = 576

Score =  242 bits (618), Expect = 1e-62
Identities = 206/654 (31%), Positives = 289/654 (43%), Gaps = 160/654 (24%)

Query: 1   MSMDASNSVESREKERDRRGRGAR-GSRFTDADGNGN-GAGSQGGGVAARDRSRERRNCR 58
           + M+   S +  + +   +  G++  +RF     + N G G + G           RN R
Sbjct: 52  VKMENDESAKEEKSDLKEKSTGSKKANRFHPYSKDKNSGTGEKKG----------PNRN-R 101

Query: 59  VYISNIPYDYRWQDLKDLFRRIVGSIEYVQLFFDESGKARGCGIVEFKDPENVQKALEKM 118
           V+ISNIPYD +WQ +KDL R  VG + YV+LF D  GK+RGCG+VEFKD E V+KALE M
Sbjct: 102 VFISNIPYDMKWQAIKDLMREKVGEVTYVELFKDAEGKSRGCGVVEFKDEEFVKKALETM 161

Query: 119 NRYEVNGRELVVKED-HGEQRDQYGRIVRDGGGGGGGGGVQGGNGGNNGGGGGGGRDHM 177
           N+Y+++GR L +KED  GE   +  +  R GG   GG                   H+
Sbjct: 162 NKYDLSGRPLNIKEDPDGENARRASQ--RTGGSFPGG----------------------HV 198

Query: 178 DDRDRGFSRRDDDRLSGRNNFNMMSNDYNNSSNYNLYGLSASFLESLGISGPLHNKVFVA 237
            D   G         L   NN N+     +N                  +G L + +FVA
Sbjct: 199 PDMGSGLMNLPPSIL---NNPNIPPEVISNLQ----------------AGRLGSTIFVA 238

Query: 238 NLDYKVDNKKLKQVFKLAGKVQSVDLSLDKEGNSRGFAVIEYDHPVEAVQAISMLDRQML 297
           NLD+KV  KKLK+VF +AG V+  D+  DK+G SRG   + ++  +EAVQAISM + Q L
Sbjct: 239 NLDFKVGWKKLKEVFSIAGTVKRADIKEDKDGKSRGMGTVTFEQAIEAVQAISMFNGQFL 298

Query: 298 FDRRMTVRLD--RIPDK------NEGIKLPEGLGGVGIGLGPNGEPLRDVAHNLPNGGQS 349
           FDR M V++D    +P +        +  +LP GLGG+G+GLGP G+P+      N+
Sbjct: 299 FDRPMHVKMDDKSVPHEEYRSHDGKTPQLPRGLGGIGMGLGPGGQPISASQLNI------ 352

Query: 350 QGQLLGNAQQGSQLGSVGSQPNSSAVSNATTNLLNNLTGVMFGNHAAVQPSPVAPVQKPS 409
            G ++GN   G                       + G  FG
Sbjct: 353 -GGVMGNLGPGGM----------------------GMDGPGFGG-------------MNR 376

Query: 410 LGNNTGSGGLNLNNLNPSILAAVVGNLGNQG--GNLSNPLLSSSL------SNLGLNLGN 461
           +G   G GGL   N          +G  G  G  G L    ++SS+        ++G+N G
Sbjct: 377 IGGGIGFGGLEAMN--------SMGGFGGVGRMGELYRGAMTSSMERDFGRGDIGINRG- 427

Query: 462 SGNDDNLPPSNVGLSNNYSSGGTGGGNSYSSGNNYSGGGGSSNLGYNAYSSS-GGMGGGN 520
                 G S     GG GG NS +         G   +G +  SSS   MG G
Sbjct: 428 ----------FGDSFGRLGGGMGGMNSVT---------GGMGMGLDRMSSSFDRMGPGI 467

Query: 521 GGVGVDGNDYNTGNPLDVYGGGSNVGNSNVGSANAVGASRKSDTIIIKNVPITCTWQTLR 580
           G +     D + G      G G       +GS         K + I  ++N+P   TWQ L+
Sbjct: 468 GAILERSIDMDRGFLSGPMGSGM---RERIGS--------KGNQIFVRNLPFDLTWQKLK 516
```

```
Query: 581 DKFREIGDVKFAEI-----RGNDVGVVRFFKERDAELAIALMDGSRLDGRNIKV 629
           +KF + G V FAEI     +    G VRF     AE A  +M+G ++ GR I V
Sbjct: 517 EKFSQCGHVMFAEIKMENGKSKGCGTVRFDSPESAEKACRIMNGIKISGREIDV 570


Score = 72.8 bits (177), Expect = 1e-11
Identities = 82/348 (23%), Positives = 133/348 (37%), Gaps = 96/348 (27%)

Query: 54  RRNCRVYISNIPYDYRWQDLKDLFRRIVGSIEYVQLFFDESGKARGCGIVEFKDPENVQK 113
           R    ++++N+ +   W+ LK++F  I G+++   + D+ GK+RG G V F+       +
Sbjct: 230 RLGSTIFVANLDFKVGWKKLKEVFS-IAGTVKRADIKEDKDGKSRGMGTVTFEQAIEAVQ 288

Query: 114 ALEKMNRYEVNGRELVVKED-----HGEQRDQYGRIVRDGGGGGGGG-------------- 155
           A+   N   + R + VK D      H E R   G+  +   G GG G
Sbjct: 289 AISMFNGQFLFDRPMHVKMDDKSVPHEEYRSHDGKTPQLPRGLGGIGMGLGPGGQPISAS 348

Query: 156 ----GGVQG-------------GNGGNNGGGGG---------------GGRDHMDDRDRGF 184
               GGV G              G GG N  GGG                 GG   M +  RG
Sbjct: 349 QLNIGGVMGNLGPGGMGMDGPGFGGMNRIGGGIGFGGLEAMNSMGGFGGVGRMGELYRGA 408

Query: 185 SRRDDDRLSGRNNFNMMS----------------NDYNNSSNYNLYGLSASFLESLG--- 225
             +R  GR +  +                     N          L  +S+SF + +G
Sbjct: 409 MTSSMERDFGRGDIGINRGFGDSFGRLGGGMGGMNSVTGGMGMGLDRMSSSF-DRMGPGI 467

Query: 226 --------------ISGPLH-----------NKVFVANLDYKVDNKKLKQVFKLAGKVQS 260
                        +SGP+             N++FV NL + +  +KLK+ F   G V
Sbjct: 468 GAILERSIDMDRGFLSGPMGSGMRERIGSKGNQIFVRNLPFDLTWQKLKEKFSQCGHVMF 527

Query: 261 VDLSLDKEGNSRGFAVIEYDHPVEAVQAISMLDRQMLFDRRMTVRLDR 308
            ++ ++  G S+G   + +D P  A +A  +++    +  R + VRLDR
Sbjct: 528 AEIKMEN-GKSKGCGTVRFDSPESAEKACRIMNGIKISGREIDVRLDR 574


Score = 68.6 bits (166), Expect = 2e-10
Identities = 41/114 (35%), Positives = 67/114 (57%), Gaps = 6/114 (5%)

Query: 20  GRGARGSRFTDADGNGNGAGSQGGGVAARDRSRERRNCRVYISNIPYDYRWQDLKDLFRR 79
           G GA   R  D D  G  +G  G G+  R+R   + N ++++ N+P+D  WQ LK+ F +
Sbjct: 466 GIGAILERSIDMD-RGFLSGPMGSGM--RERIGSKGN-QIFVRNLPFDLTWQKLKEKFSQ 521

Query: 80  IVGSIEYVQLFFDESGKARGCGIVEFKDPENVQKALEKMNRYEVNGRELVVKED 133
             G + + ++   E+GK++GCG V F   PE+ +KA    N   +++GRE+ V+ D
Sbjct: 522 C-GHVMFAEIKM-ENGKSKGCGTVRFDSPESAEKACRIMNGIKISGREIDVRLD 573
```

Homology to human protein NP057216.1 (GenBank Accession Number)

```
ref|NP_057216.1| (NM_016132) myelin gene expression factor 2 [Homo sapiens]
gb|AAD43038.1| (AF106685) myelin gene expression factor 2 [Homo sapiens]
Length = 547

 Score =  238 bits (607), Expect = 2e-61
 Identities = 204/659 (30%), Positives = 295/659 (43%), Gaps = 150/659 (22%)

Query: 3   MDASNSVESREKERDRRGRGAR-GSRFTDADGNGN-GAGSQGGGVAARDRSRERRNCRVY 60
           M+   S +  + +   + G++  +RF     + N G G + G           RN RV+
Sbjct: 1   MENDESAKEEKSDLKEKSTGSKKANRFHPYSKDKNSGTGEKKG----------PNRN-RVF 50

Query: 61  ISNIPYDYRWQDLKDLFRRIVGSIEYVQLFFDESGKARGCGIVEFKDPENVQKALEKMNR 120
           ISNIPYD +WQ +KDL R  VG + YV+LF D  GK+RGCG+VEFKD E V+KALE MN+
Sbjct: 51  ISNIPYDMKWQAIKDLMREKVGEVTYVELFKDAEGKSRGCGVVEFKDEEFVKKALETMNK 110

Query: 121 YEVNGRELVVKEDHGEQRDQYGRIVRDGGGGGGGGGVQGGNGGNNGGGGGGGRDHMDDR 180
```

```
           Y+++GR + +KED   +  +   + R G          QG +  + G G
Sbjct: 111 YDLSGRRVNIKEDPDGENARRA-LQRTGTS-------FQGSHASDVGSG----------- 151

Query: 181 DRGFSRRDDDRLSGRNNFNMMSNDYNNSSNYNLYGLSASFLESLGISGPLHNKVFVANLD 240
                             N+  +  NN +      +    + +L  +G L + +FVANLD
Sbjct: 152 ----------------LVNLPPSILNNPN------IPPEVISNLQ-AGRLGSTIFVANLD 188

Query: 241 YKVDNKKLKQVFKLAGKVQSVDLSLDKEGNSRGFAVIEYDHPVEAVQAISMLDRQMLFDR 300
           +KV  KKLK+VF +AG V++     DK+G SRG   + ++  +EAVQAISM + Q LFDR
Sbjct: 189 FKVGWKKLKEVFSIAGTVKAGSYKEDKDGKSRGMGTVTFEQAIEAVQAISMFNGQFLFDR 248

Query: 301 RMTVRLD---------RIPDKNEGIKLPEGLGGVGIGLGPNGEPLRDVAHNLPNGGQSQG 351
            M V++D         R PD  +  +LP GLGG+G+GLGP G+P+     N+       G
Sbjct: 249 PMHVKMDDKSVPHEEYRSPD-GKTPQLPRGLGGIGMGLGPGGQPISASQLNI-------G 300

Query: 352 QLLGNAQQGSQLGSVGSQPNSSAVSNATTNLLNNLTGVMFGNHAAVQPSPVAPVQKPSLG 411
            ++GN   G                         + G  FG                 +G
Sbjct: 301 GVMGNLGPGGM----------------------GMDGPGFGG-------------MNRIG 325

Query: 412 NNTGSGGLNLNNLNPSILAAVVGNLGNQG--GNLSNPLLSSSLS------NLGLNLGNSG 463
              G GGL   N         +G  G  G  G L    ++SS+       ++GL+ G
Sbjct: 326 GGIGFGGLEAMN--------SMGGFGGVGRMGELYRGAMTSSMERDFGHRDIGLSRGFGD 377

Query: 464 NDDNLPPSNVGLSNNYSSGGTGGGNSYSSGNNYSGGGGSSN-------LGYNAYSSS-GG 515
           +   L  + +G      +G  G  N    G+  SGG GS N       +G +  SSS
Sbjct: 378 SFGRLGSAMIG----GITGRIGSSNMGPVGSGISGGMGSMNSVTGGMGMGLDRMSSSFDR 433

Query: 516 MGGGNGGVGVDGNDYNTGNPLDVYGGGSNVGNSNVGSANAVGASRKSDTIIIKNVPITCT 575
           MG G G +     D + G      G G       +GS        K + I ++N+P   T
Sbjct: 434 MGPGIGAILERSIDMDRGFLSGPMGSGM---RERIGS--------KGNQIFVRNLPFDLT 482

Query: 576 WQTLRDKFREIGDVKFAEI-----RGNDVGVVRFFKERDAELAIALMDGSRLDGRNIKV 629
           WQ L++KF + G V FAEI     +    G VRF     AE A  +M+G ++ GR I V
Sbjct: 483 WQKLKEKFSQCGHVMFAEIKMENGKSKGCGTVRFDSPESAEKACRIMNGIKISGREIDV 541


Score = 68.6 bits (166), Expect = 2e-10
Identities = 41/114 (35%), Positives = 67/114 (57%), Gaps = 6/114 (5%)

Query: 20  GRGARGSRFTDADGNGNGAGSQGGGVAARDRSRERRNCRVYISNIPYDYRWQDLKDLFRR 79
           G GA   R  D D  G  +G  G G+  R+R   + N ++++ N+P+D  WQ LK+ F +
Sbjct: 437 GIGAILERSIDMD-RGFLSGPMGSGM--RERIGSKGN-QIFVRNLPFDLTWQKLKEKFSQ 492

Query: 80  IVGSIEYVQLFFDESGKARGCGIVEFKDPENVQKALEKMNRYEVNGRELVVKED 133
             G + + ++   E+GK++GCG V F  PE+ +KA   MN  +++GRE+ V+ D
Sbjct: 493 C-GHVMFAEIKM-ENGKSKGCGTVRFDSPESAEKACRIMNGIKISGREIDVRLD 544


Score = 55.5 bits (132), Expect = 2e-06
Identities = 41/157 (26%), Positives = 69/157 (43%), Gaps = 11/157 (7%)

Query: 152 GGGGGGVQGGNGGNNGGGGGGGRDHMDDRDRGFSRRDDDRLSGRNNFNMMSNDYNNSSNY 211
           G  G G+ GG G  N   GG G   +D     F R       G     ++    +    +
Sbjct: 400 GPVGSGISGGMGSMNSVTGGMGMG-LDRMSSSFDRM------GPGIGAILERSIDMDRGF 452

Query: 212 NLYGLSASFLESLGISGPLHNKVFVANLDYKVDNKKLKQVFKLAGKVQSVDLSLDKEGNS 271
               + +   E +G  G   N++FV NL + +  +KLK+ F   G V   ++ ++  G S
Sbjct: 453 LSGPMGSGMRERIGSKG---NQIFVRNLPFDLTWQKLKEKFSQCGHVMFAEIKMEN-GKS 508

Query: 272 RGFAVIEYDHPVEAVQAISMLDRQMLFDRRMTVRLDR 308
           +G   + +D P    +A  +++   +  R + VRLDR
Sbjct: 509 KGCGTVRFDSPESAEKACRIMNGIKISGREIDVRLDR 545
```

FIGURE 10. Triglyceride content of a Drosophila *how* (GadFly Accession Number CG10293) mutant

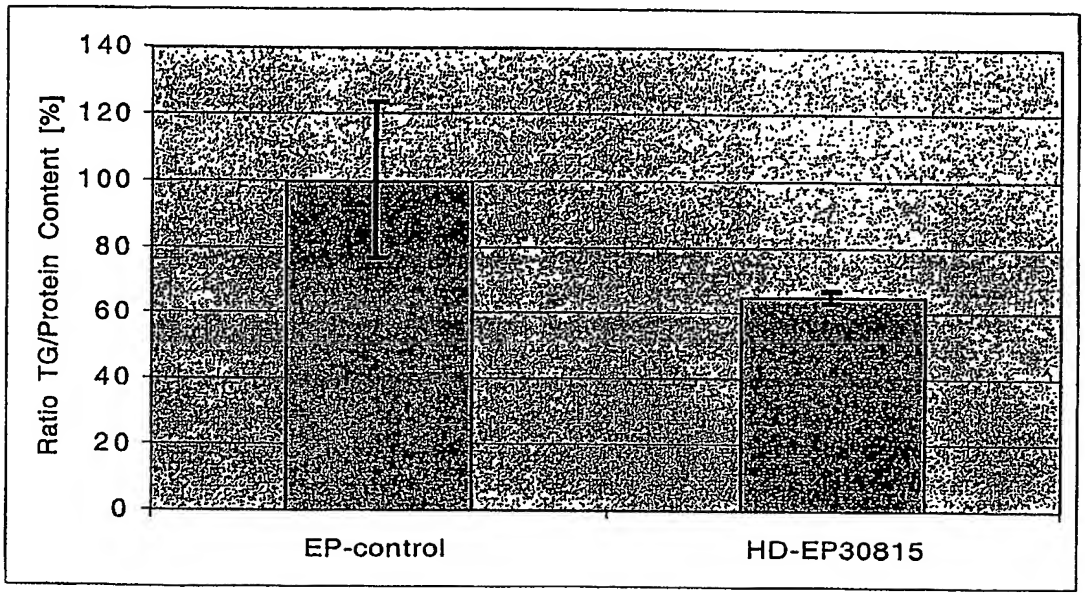

FIGURE 11. Molecular organisation of the *how* gene (GadFly Accession Number CG10293)

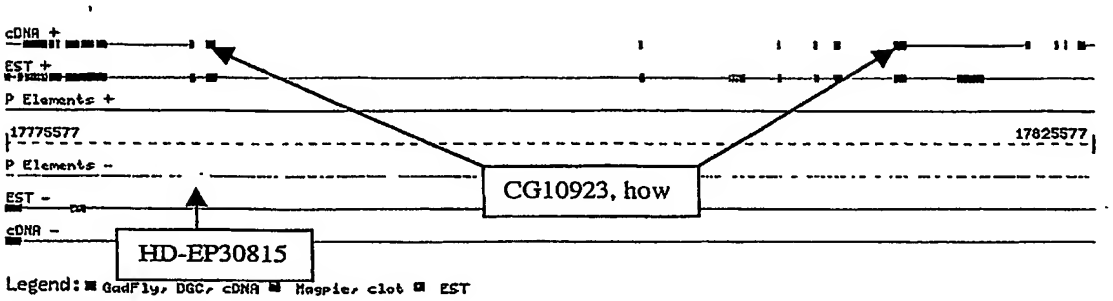

FIGURE 12. BLASTP results for CG10293 (GadFly Accession Number)

```
gb|AAF63416.1|AF142421_1 (AF142421) QUAKING isoform 5 [Homo sapiens]
Length = 337

Score =  289 bits (739), Expect = 5e-77
Identities = 168/334 (50%), Positives = 215/334 (64%), Gaps = 20/334 (5%)

Query: 67  QQQQSTQSIADYLAQLLKDRKQLAAFPN---VFTHVERLLDEEIARVRASLFQ--ING-V 120
           + ++  +    DYL QL+ D+K +++ PN    +F H+ERLLDEEI+RVR  ++    +NG
Sbjct: 2   ETKEKPKPTPDYLMQLMNDKKLMSSLPNFCGIFNHLERLLDEEISRVRKDMYNDTLNGST 61

Query: 121 KKEPLTLPEPEGSVVTMNEKVYVPVREHPDFNFVGRILGPRGMTAKQLEQETGCKIMVRG 180
           +K    LP+  G +V + EK+YVPV+E+PDFNFVGRILGPRG+TAKQLE ETGCKIMVRG
Sbjct: 62  EKRSAELPDAVGPIVQLQEKLYVPVKEYPDFNFVGRILGPRGLTAKQLEAETGCKIMVRG 121

Query: 181 KGSMRDKKKEDANRGKPNWEHLSDDLHVLITVEDTENRATVKLAQAVAEVQKLLVPQAEG 240
           KGSMRDKKKE+ NRGKPNWEHL++DLHVLITVED +NRA +KL +AV EV+KLLVP AEG
Sbjct: 122 KGSMRDKKKEEQNRGKPNWEHLNEDLHVLITVEDAQNRAEIKLKRAVEEVKKLLVPAAEG 181

Query: 241 EDELKKRQLMELAIINGTYRDTTAKSVAVCDEEWRRLVAASDSRLLTSTGLPGLAAQIRA 300
           ED LKK QLMELAI+NGTYRD   KS A+          A +  R++T         A +R
Sbjct: 182 EDSLKKMQLMELAILNGTYRDANIKSPALAFS--LAATAQAAPRIITGPAPVLPPAALRT 239

Query: 301 PA-AAPLGAPLILNPRMTVPTTAASILSAQAAPTAAFDQTG--HGMIFAPYDYANYAALA 357
           P  A P    PLI    +  V       + +     PTAA     G    G+I+ PY+Y  Y
Sbjct: 240 PTPAGPTIMPLIRQIQTAV------MPNGTPHPTAAIVPPGPEAGLIYTPYEYP-YTLAP 292

Query: 358 GNPLLTEYADHS--VGAIKQQRRLATNREHPYQR 389
              +L    + S  +GA+  + R    R HPYQR
Sbjct: 293 ATSILEYPIEPSGVLGAVATKVRRHDMRVHPYQR 326


ref|XP_037438.2| (XM_037438) similar to KH domain RNA binding protein QKI-5A [Homo
sapiens]
Length = 341

Score =  289 bits (739), Expect = 5e-77
Identities = 168/334 (50%), Positives = 215/334 (64%), Gaps = 20/334 (5%)

Query: 67  QQQQSTQSIADYLAQLLKDRKQLAAFPN---VFTHVERLLDEEIARVRASLFQ--ING-V 120
           + ++  +    DYL QL+ D+K +++ PN    +F H+ERLLDEEI+RVR  ++    +NG
Sbjct: 6   ETKEKPKPTPDYLMQLMNDKKLMSSLPNFCGIFNHLERLLDEEISRVRKDMYNDTLNGST 65

Query: 121 KKEPLTLPEPEGSVVTMNEKVYVPVREHPDFNFVGRILGPRGMTAKQLEQETGCKIMVRG 180
           +K    LP+  G +V + EK+YVPV+E+PDFNFVGRILGPRG+TAKQLE ETGCKIMVRG
Sbjct: 66  EKRSAELPDAVGPIVQLQEKLYVPVKEYPDFNFVGRILGPRGLTAKQLEAETGCKIMVRG 125

Query: 181 KGSMRDKKKEDANRGKPNWEHLSDDLHVLITVEDTENRATVKLAQAVAEVQKLLVPQAEG 240
           KGSMRDKKKE+ NRGKPNWEHL++DLHVLITVED +NRA +KL +AV EV+KLLVP AEG
Sbjct: 126 KGSMRDKKKEEQNRGKPNWEHLNEDLHVLITVEDAQNRAEIKLKRAVEEVKKLLVPAAEG 185

Query: 241 EDELKKRQLMELAIINGTYRDTTAKSVAVCDEEWRRLVAASDSRLLTSTGLPGLAAQIRA 300
           ED LKK QLMELAI+NGTYRD   KS A+          A +  R++T         A +R
Sbjct: 186 EDSLKKMQLMELAILNGTYRDANIKSPALAFS--LAATAQAAPRIITGPAPVLPPAALRT 243

Query: 301 PA-AAPLGAPLILNPRMTVPTTAASILSAQAAPTAAFDQTG--HGMIFAPYDYANYAALA 357
           P  A P    PLI    +  V       + +     PTAA     G    G+I+ PY+Y  Y
Sbjct: 244 PTPAGPTIMPLIRQIQTAV------MPNGTPHPTAAIVPPGPEAGLIYTPYEYP-YTLAP 296

Query: 358 GNPLLTEYADHS--VGAIKQQRRLATNREHPYQR 389
              +L    + S  +GA+  + R    R HPYQR
Sbjct: 297 ATSILEYPIEPSGVLGAVATKVRRHDMRVHPYQR 330
```

gb|AAF63414.1|AF142419_1 (AF142419) QUAKING isoform 6 [Homo sapiens]
Length = 363

Score = 289 bits (739), Expect = 5e-77
Identities = 168/334 (50%), Positives = 215/334 (64%), Gaps = 20/334 (5%)

```
Query: 67  QQQQSTQSIADYLAQLLKDRKQLAAFPN---VFTHVERLLDEEIARVRASLFQ--ING-V 120
           + ++  +   DYL QL+ D+K +++ PN   +F H+ERLLDEEI+RVR  ++   +NG
Sbjct: 28  ETKEKPKPTPDYLMQLMNDKKLMSSLPNFCGIFNHLERLLDEEISRVRKDMYNDTLNGST 87

Query: 121 KKEPLTLPEPEGSVVTMNEKVYVPVREHPDFNFVGRILGPRGMTAKQLEQETGCKIMVRG 180
           +K    LP+  G +V + EK+YVPV+E+PDFNFVGRILGPRG+TAKQLE ETGCKIMVRG
Sbjct: 88  EKRSAELPDAVGPIVQLQEKLYVPVKEYPDFNFVGRILGPRGLTAKQLEAETGCKIMVRG 147

Query: 181 KGSMRDKKKEDANRGKPNWEHLSDDLHVLITVEDTENRATVKLAQAVAEVQKLLVPQAEG 240
           KGSMRDKKKE+ NRGKPNWEHL++DLHVLITVED +NRA +KL +AV EV+KLLVP AEG
Sbjct: 148 KGSMRDKKKEEQNRGKPNWEHLNEDLHVLITVEDAQNRAEIKLKRAVEEVKKLLVPAAEG 207

Query: 241 EDELKKRQLMELAIINGTYRDTTAKSVAVCDEEWRRLVAASDSRLLTSTGLPGLAAQIRA 300
           ED LKK QLMELAI+NGTYRD   KS A+          A +  R++T         A +R
Sbjct: 208 EDSLKKMQLMELAILNGTYRDANIKSPALAFS--LAATAQAAPRIITGPAPVLPPAALRT 265

Query: 301 PA-AAPLGAPLILNPRMTVPTTAASILSAQAAPTAAFDQTG--HGMIFAPYDYANYAALA 357
           P  A P   PLI   +  V      + +    PTAA    G   G+I+ PY+Y  Y
Sbjct: 266 PTPAGPTIMPLIRQIQTAV------MPNGTPHPTAAIVPPGPEAGLIYTPYEYP-YTLAP 318

Query: 358 GNPLLTEYADHS--VGAIKQQRRLATNREHPYQR 389
              +L    + S  +GA+  + R    R HPYQR
Sbjct: 319 ATSILEYPIEPSGVLGAVATKVRRHDMRVHPYQR 352
```

dbj|BAB55032.1| (AK027309) unnamed protein product [Homo sapiens]
Length = 323

Score = 282 bits (722), Expect = 5e-75
Identities = 165/320 (51%), Positives = 208/320 (64%), Gaps = 20/320 (6%)

```
Query: 81  QLLKDRKQLAAFPN---VFTHVERLLDEEIARVRASLFQ--ING-VKKEPLTLPEPEGSV 134
           QL+ D+K +++ PN   +F H+ERLLDEEI+RVR  ++   +NG  +K    LP+  G +
Sbjct: 2   QLMNDKKLMSSLPNFCGIFNHLERLLDEEISRVRKDMYNDTLNGSTEKRSAELPDAVGPI 61

Query: 135 VTMNEKVYVPVREHPDFNFVGRILGPRGMTAKQLEQETGCKIMVRGKGSMRDKKKEDANR 194
           V + EK+YVPV+E+PDFNFVGRILGPRG+TAKQLE ETGCKIMVRGKGSMRDKKKE+ NR
Sbjct: 62  VQLQEKLYVPVKEYPDFNFVGRILGPRGLTAKQLEAETGCKIMVRGKGSMRDKKKEEQNR 121

Query: 195 GKPNWEHLSDDLHVLITVEDTENRATVKLAQAVAEVQKLLVPQAEGEDELKKRQLMELAI 254
           GKPNWEHL++DLHVLITVED +NRA +KL +AV EV+KLLVP AEGED LKK QLMELAI
Sbjct: 122 GKPNWEHLNEDLHVLITVEDAQNRAEIKLKRAVEEVKKLLVPAAEGEDSLKKMQLMELAI 181

Query: 255 INGTYRDTTAKSVAVCDEEWRRLVAASDSRLLTSTGLPGLAAQIRAPA-AAPLGAPLILN 313
           +NGTYRD   KS A+          A +  R++T         A +R P  A P   PLI
Sbjct: 182 LNGTYRDANIKSPALAFS--LAATAQAAPRIITGPAPVLPPAALRTPTPAGPTIMPLIRQ 239

Query: 314 PRMTVPTTAASILSAQAAPTAAFDQTG--HGMIFAPYDYANYAALAGNPLLTEYADHS-- 369
            +  V      + +    PTAA    G   G+I+ PY+Y  Y       +L    + S
Sbjct: 240 IQTAV------MPNGTPHPTAAIVPPGPEAGLIYTPYEYP-YTLAPATSILEYPIEPSGV 292

Query: 370 VGAIKQQRRLATNREHPYQR 389
           +GA+  + R    R HPYQR
Sbjct: 293 LGAVATKVRRHDMRVHPYQR 312
```

gb|AAF63413.1|AF142418_1 (AF142418) QUAKING isoform 2 [Homo sapiens]
Length = 347

Score = 280 bits (716), Expect = 2e-74
Identities = 156/293 (53%), Positives = 198/293 (67%), Gaps = 17/293 (5%)

```
Query: 67  QQQQSTQSIADYLAQLLKDRKQLAAFPN---VFTHVERLLDEEIARVRASLFQ--ING-V 120
           + ++  +   DYL QL+ D+K +++ PN   +F H+ERLLDEEI+RVR  ++   +NG
Sbjct: 28  ETKEKPKPTPDYLMQLMNDKKLMSSLPNFCGIFNHLERLLDEEISRVRKDMYNDTLNGST 87

Query: 121 KKEPLTLPEPEGSVVTMNEKVYVPVREHPDFNFVGRILGPRGMTAKQLEQETGCKIMVRG 180
           +K    LP+  G +V + EK+YVPV+E+PDFNFVGRILGPRG+TAKQLE ETGCKIMVRG
Sbjct: 88  EKRSAELPDAVGPIVQLQEKLYVPVKEYPDFNFVGRILGPRGLTAKQLEAETGCKIMVRG 147

Query: 181 KGSMRDKKKEDANRGKPNWEHLSDDLHVLITVEDTENRATVKLAQAVAEVQKLLVPQAEG 240
           KGSMRDKKKE+ NRGKPNWEHL++DLHVLITVED +NRA +KL +AV EV+KLLVP AEG
Sbjct: 148 KGSMRDKKKEEQNRGKPNWEHLNEDLHVLITVEDAQNRAEIKLKRAVEEVKKLLVPAAEG 207

Query: 241 EDELKKRQLMELAIINGTYRDTTAKSVAVCDEEWRRLVAASDSRLLTSTGLPGLAAQIRA 300
           ED LKK QLMELAI+NGTYRD   KS A+          A +  R++T         A +R
Sbjct: 208 EDSLKKMQLMELAILNGTYRDANIKSPALAFS--LAATAQAAPRIITGPAPVLPPAALRT 265

Query: 301 PA-AAPLGAPLILNPRMTVPTTAASILSAQAAPTAAFDQTG--HGMIFAPYDY 350
           P  A P   PLI   +  V      + +    PTAA    G   G+I+ PY+Y
Sbjct: 266 PTPAGPTIMPLIRQIQTAV------MPNGTPHPTAAIVPPGPEAGLIYTPYEY 312
```

gb|AAF63417.1|AF142422_1 (AF142422) QUAKING isoform 3 [Homo sapiens]
Length = 341

Score = 280 bits (716), Expect = 2e-74
Identities = 156/293 (53%), Positives = 198/293 (67%), Gaps = 17/293 (5%)

```
Query: 67  QQQQSTQSIADYLAQLLKDRKQLAAFPN---VFTHVERLLDEEIARVRASLFQ--ING-V 120
           + ++  +   DYL QL+ D+K +++ PN   +F H+ERLLDEEI+RVR  ++   +NG
Sbjct: 28  ETKEKPKPTPDYLMQLMNDKKLMSSLPNFCGIFNHLERLLDEEISRVRKDMYNDTLNGST 87

Query: 121 KKEPLTLPEPEGSVVTMNEKVYVPVREHPDFNFVGRILGPRGMTAKQLEQETGCKIMVRG 180
           +K    LP+  G +V + EK+YVPV+E+PDFNFVGRILGPRG+TAKQLE ETGCKIMVRG
Sbjct: 88  EKRSAELPDAVGPIVQLQEKLYVPVKEYPDFNFVGRILGPRGLTAKQLEAETGCKIMVRG 147

Query: 181 KGSMRDKKKEDANRGKPNWEHLSDDLHVLITVEDTENRATVKLAQAVAEVQKLLVPQAEG 240
           KGSMRDKKKE+ NRGKPNWEHL++DLHVLITVED +NRA +KL +AV EV+KLLVP AEG
Sbjct: 148 KGSMRDKKKEEQNRGKPNWEHLNEDLHVLITVEDAQNRAEIKLKRAVEEVKKLLVPAAEG 207

Query: 241 EDELKKRQLMELAIINGTYRDTTAKSVAVCDEEWRRLVAASDSRLLTSTGLPGLAAQIRA 300
           ED LKK QLMELAI+NGTYRD   KS A+          A +  R++T         A +R
Sbjct: 208 EDSLKKMQLMELAILNGTYRDANIKSPALAFS--LAATAQAAPRIITGPAPVLPPAALRT 265

Query: 301 PA-AAPLGAPLILNPRMTVPTTAASILSAQAAPTAAFDQTG--HGMIFAPYDY 350
           P  A P   PLI   +  V      + +    PTAA    G   G+I+ PY+Y
Sbjct: 266 PTPAGPTIMPLIRQIQTAV------MPNGTPHPTAAIVPPGPEAGLIYTPYEY 312
```

gb|AAF63415.1|AF142420_1 (AF142420) QUAKING isoform 4 [Homo sapiens]
Length = 315

Score = 280 bits (716), Expect = 2e-74
Identities = 156/293 (53%), Positives = 198/293 (67%), Gaps = 17/293 (5%)

```
Query: 67  QQQQSTQSIADYLAQLLKDRKQLAAFPN---VFTHVERLLDEEIARVRASLFQ--ING-V 120
           + ++  +   DYL QL+ D+K +++ PN   +F H+ERLLDEEI+RVR  ++   +NG
Sbjct: 2   ETKEKPKPTPDYLMQLMNDKKLMSSLPNFCGIFNHLERLLDEEISRVRKDMYNDTLNGST 61
```

```
Query: 121 KKEPLTLPEPEGSVVTMNEKVYVPVREHPDFNFVGRILGPRGMTAKQLEQETGCKIMVRG 180
           +K    LP+  G +V + EK+YVPV+E+PDFNFVGRILGPRG+TAKQLE ETGCKIMVRG
Sbjct: 62  EKRSAELPDAVGPIVQLQEKLYVPVKEYPDFNFVGRILGPRGLTAKQLEAETGCKIMVRG 121

Query: 181 KGSMRDKKKEDANRGKPNWEHLSDDLHVLITVEDTENRATVKLAQAVAEVQKLLVPQAEG 240
           KGSMRDKKKE+ NRGKPNWEHL++DLHVLITVED +NRA +KL +AV EV+KLLVP AEG
Sbjct: 122 KGSMRDKKKEEQNRGKPNWEHLNEDLHVLITVEDAQNRAEIKLKRAVEEVKKLLVPAAEG 181

Query: 241 EDELKKRQLMELAIINGTYRDTTAKSVAVCDEEWRRLVAASDSRLLTSTGLPGLAAQIRA 300
           ED LKK QLMELAI+NGTYRD   KS A+          A +  R++T         A +R
Sbjct: 182 EDSLKKMQLMELAILNGTYRDANIKSPALAFS--LAATAQAAPRIITGPAPVLPPAALRT 239

Query: 301 PA-AAPLGAPLILNPRMTVPTTAASILSAQAAPTAAFDQTG--HGMIFAPYDY 350
           P  A P   PLI   +  V      + +    PTAA    G   G+I+ PY+Y
Sbjct: 240 PTPAGPTIMPLIRQIQTAV------MPNGTPHPTAAIVPPGPEAGLIYTPYEY 286


gb|AAC99453.1| (AF090403) KH domain RNA binding protein QKI-5B [Mus musculus]
gb|AAC99454.1| (AF090404) KH domain RNA binding protein QKI-6 [Mus musculus]
gb|AAD53330.1| (AF090401) QKI-6 protein [Mus musculus]
dbj|BAB69497.1| (AB067799) RNA binding protein HQK-6 [Homo sapiens]
Length = 319

Score =  280 bits (716), Expect = 2e-74
Identities = 156/293 (53%), Positives = 198/293 (67%), Gaps = 17/293 (5%)

Query: 67  QQQQSTQSIADYLAQLLKDRKQLAAFPN---VFTHVERLLDEEIARVRASLFQ--ING-V 120
           + ++  +   DYL QL+ D+K +++ PN   +F H+ERLLDEEI+RVR  ++   +NG
Sbjct: 6   ETKEKPKPTPDYLMQLMNDKKLMSSLPNFCGIFNHLERLLDEEISRVRKDMYNDTLNGST 65

Query: 121 KKEPLTLPEPEGSVVTMNEKVYVPVREHPDFNFVGRILGPRGMTAKQLEQETGCKIMVRG 180
           +K    LP+  G +V + EK+YVPV+E+PDFNFVGRILGPRG+TAKQLE ETGCKIMVRG
Sbjct: 66  EKRSAELPDAVGPIVQLQEKLYVPVKEYPDFNFVGRILGPRGLTAKQLEAETGCKIMVRG 125

Query: 181 KGSMRDKKKEDANRGKPNWEHLSDDLHVLITVEDTENRATVKLAQAVAEVQKLLVPQAEG 240
           KGSMRDKKKE+ NRGKPNWEHL++DLHVLITVED +NRA +KL +AV EV+KLLVP AEG
Sbjct: 126 KGSMRDKKKEEQNRGKPNWEHLNEDLHVLITVEDAQNRAEIKLKRAVEEVKKLLVPAAEG 185

Query: 241 EDELKKRQLMELAIINGTYRDTTAKSVAVCDEEWRRLVAASDSRLLTSTGLPGLAAQIRA 300
           ED LKK QLMELAI+NGTYRD   KS A+          A +  R++T         A +R
Sbjct: 186 EDSLKKMQLMELAILNGTYRDANIKSPALAFS--LAATAQAAPRIITGPAPVLPPAALRT 243

Query: 301 PA-AAPLGAPLILNPRMTVPTTAASILSAQAAPTAAFDQTG--HGMIFAPYDY 350
           P  A P   PLI   +  V      + +    PTAA    G   G+I+ PY+Y
Sbjct: 244 PTPAGPTIMPLIRQIQTAV------MPNGTPHPTAAIVPPGPEAGLIYTPYEY 290


dbj|BAB69499.1| (AB067801) RNA binding protein HQK-7B [Homo sapiens]
Length = 319

Score =  280 bits (716), Expect = 2e-74
Identities = 156/293 (53%), Positives = 198/293 (67%), Gaps = 17/293 (5%)

Query: 67  QQQQSTQSIADYLAQLLKDRKQLAAFPN---VFTHVERLLDEEIARVRASLFQ--ING-V 120
           + ++  +   DYL QL+ D+K +++ PN   +F H+ERLLDEEI+RVR  ++   +NG
Sbjct: 6   ETKEKPKPTPDYLMQLMNDKKLMSSLPNFCGIFNHLERLLDEEISRVRKDMYNDTLNGST 65

Query: 121 KKEPLTLPEPEGSVVTMNEKVYVPVREHPDFNFVGRILGPRGMTAKQLEQETGCKIMVRG 180
           +K    LP+  G +V + EK+YVPV+E+PDFNFVGRILGPRG+TAKQLE ETGCKIMVRG
Sbjct: 66  EKRSAELPDAVGPIVQLQEKLYVPVKEYPDFNFVGRILGPRGLTAKQLEAETGCKIMVRG 125

Query: 181 KGSMRDKKKEDANRGKPNWEHLSDDLHVLITVEDTENRATVKLAQAVAEVQKLLVPQAEG 240
           KGSMRDKKKE+ NRGKPNWEHL++DLHVLITVED +NRA +KL +AV EV+KLLVP AEG
Sbjct: 126 KGSMRDKKKEEQNRGKPNWEHLNEDLHVLITVEDAQNRAEIKLKRAVEEVKKLLVPAAEG 185
```

```
Query: 241 EDELKKRQLMELAIINGTYRDTTAKSVAVCDEEWRRLVAASDSRLLTSTGLPGLAAQIRA 300
           ED LKK QLMELAI+NGTYRD   KS A+          A +  R++T         A +R
Sbjct: 186 EDSLKKMQLMELAILNGTYRDANIKSPALAFS--LAATAQAAPRIITGPAPVLPPAALRT 243

Query: 301 PA-AAPLGAPLILNPRMTVPTTAASILSAQAAPTAAFDQTG--HGMIFAPYDY 350
           P  A P    PLI    +  V       + +    PTAA    G   G+I+ PY+Y
Sbjct: 244 PTPAGPTIMPLIRQIQTAV------MPNGTPHPTAAIVPPGPEAGLIYTPYEY 290


ref|NP_068681.1| (NM_021881) quaking protein [Mus musculus]
gb|AAC52491.1| (U44940) qkI-7 [Mus musculus]
gb|AAD53331.1| (AF090401) QKI-7 protein [Mus musculus]
dbj|BAB69498.1| (AB067800) RNA binding protein HQK-7 [Homo sapiens]
dbj|BAB69681.1| (AB067809) RNA binding protein HQK [Homo sapiens]
prf||2208447A RNA-binding/signal transduction protein:ISOTYPE=I [Mus musculus]
Length = 325

Score =  280 bits (716), Expect = 2e-74
Identities = 156/293 (53%), Positives = 198/293 (67%), Gaps = 17/293 (5%)

Query: 67  QQQQSTQSIADYLAQLLKDRKQLAAFPN---VFTHVERLLDEEIARVRASLFQ--ING-V 120
           + ++  +    DYL QL+ D+K +++ PN    +F H+ERLLDEEI+RVR  ++    +NG
Sbjct: 6   ETKEKPKPTPDYLMQLMNDKKLMSSLPNFCGIFNHLERLLDEEISRVRKDMYNDTLNGST 65

Query: 121 KKEPLTLPEPEGSVVTMNEKVYVPVREHPDFNFVGRILGPRGMTAKQLEQETGCKIMVRG 180
           +K    LP+  G +V + EK+YVPV+E+PDFNFVGRILGPRG+TAKQLE ETGCKIMVRG
Sbjct: 66  EKRSAELPDAVGPIVQLQEKLYVPVKEYPDFNFVGRILGPRGLTAKQLEAETGCKIMVRG 125

Query: 181 KGSMRDKKKEDANRGKPNWEHLSDDLHVLITVEDTENRATVKLAQAVAEVQKLLVPQAEG 240
           KGSMRDKKKE+ NRGKPNWEHL++DLHVLITVED +NRA +KL +AV EV+KLLVP AEG
Sbjct: 126 KGSMRDKKKEEQNRGKPNWEHLNEDLHVLITVEDAQNRAEIKLKRAVEEVKKLLVPAAEG 185

Query: 241 EDELKKRQLMELAIINGTYRDTTAKSVAVCDEEWRRLVAASDSRLLTSTGLPGLAAQIRA 300
           ED LKK QLMELAI+NGTYRD   KS A+          A +  R++T         A +R
Sbjct: 186 EDSLKKMQLMELAILNGTYRDANIKSPALAFS--LAATAQAAPRIITGPAPVLPPAALRT 243

Query: 301 PA-AAPLGAPLILNPRMTVPTTAASILSAQAAPTAAFDQTG--HGMIFAPYDY 350
           P  A P    PLI    +  V       + +    PTAA    G   G+I+ PY+Y
Sbjct: 244 PTPAGPTIMPLIRQIQTAV------MPNGTPHPTAAIVPPGPEAGLIYTPYEY 290


gb|AAF63412.1|AF142417_1 (AF142417) QUAKING isoform 1 [Homo sapiens]
Length = 321

Score =  280 bits (716), Expect = 2e-74
Identities = 156/293 (53%), Positives = 198/293 (67%), Gaps = 17/293 (5%)

Query: 67  QQQQSTQSIADYLAQLLKDRKQLAAFPN---VFTHVERLLDEEIARVRASLFQ--ING-V 120
           + ++  +    DYL QL+ D+K +++ PN    +F H+ERLLDEEI+RVR  ++    +NG
Sbjct: 2   ETKEKPKPTPDYLMQLMNDKKLMSSLPNFCGIFNHLERLLDEEISRVRKDMYNDTLNGST 61

Query: 121 KKEPLTLPEPEGSVVTMNEKVYVPVREHPDFNFVGRILGPRGMTAKQLEQETGCKIMVRG 180
           +K    LP+  G +V + EK+YVPV+E+PDFNFVGRILGPRG+TAKQLE ETGCKIMVRG
Sbjct: 62  EKRSAELPDAVGPIVQLQEKLYVPVKEYPDFNFVGRILGPRGLTAKQLEAETGCKIMVRG 121

Query: 181 KGSMRDKKKEDANRGKPNWEHLSDDLHVLITVEDTENRATVKLAQAVAEVQKLLVPQAEG 240
           KGSMRDKKKE+ NRGKPNWEHL++DLHVLITVED +NRA +KL +AV EV+KLLVP AEG
Sbjct: 122 KGSMRDKKKEEQNRGKPNWEHLNEDLHVLITVEDAQNRAEIKLKRAVEEVKKLLVPAAEG 181

Query: 241 EDELKKRQLMELAIINGTYRDTTAKSVAVCDEEWRRLVAASDSRLLTSTGLPGLAAQIRA 300
           ED LKK QLMELAI+NGTYRD   KS A+          A +  R++T         A +R
Sbjct: 182 EDSLKKMQLMELAILNGTYRDANIKSPALAFS--LAATAQAAPRIITGPAPVLPPAALRT 239
```

```
Query: 301 PA-AAPLGAPLILNPRMTVPTTAASILSAQAAPTAAFDQTG--HGMIFAPYDY 350
          P  A P    PLI    +  V       + +     PTAA    G    G+I+ PY+Y
Sbjct: 240 PTPAGPTIMPLIRQIQTAV------MPNGTPHPTAAIVPPGPEAGLIYTPYEY 286


dbj|BD004960.1| Genes related to stomach cancer
Length = 1993

Score =  288 bits (738), Expect = 1e-77
Identities = 168/324 (51%), Positives = 211/324 (64%), Gaps = 11/324 (3%)
Frame = +1

Query: 77  DYLAQLLKDRKQLAAFPN---VFTHVERLLDEEIARVRASLFQ--ING-VKKEPLTLPEP 130
           DYL QL+ D+K +++ PN    +F H+ERLLDEEI+RVR  ++    +NG  +K     LP+
Sbjct: 4   DYLMQLMNDKKLMSSLPNFCGIFNHLERLLDEEISRVRKDMYNDTLNGSTEKRSAELPDA 183

Query: 131 EGSVVTMNEKVYVPVREHPDFNFVGRILGPRGMTAKQLEQETGCKIMVRGKGSMRDKKKE 190
            G +V + EK+YVPV+E+PDFNFVGRILGPRG+TAKQLE ETGCKIMVRGKGSMRDKKKE
Sbjct: 184 VGPIVQLQEKLYVPVKEYPDFNFVGRILGPRGLTAKQLEAETGCKIMVRGKGSMRDKKKE 363

Query: 191 DANRGKPNWEHLSDDLHVLITVEDTENRATVKLAQAVAEVQKLLVPQAEGEDELKKRQLM 250
           + NRGKPNWEHL++DLHVLITVED +NRA +KL +AV EV+KLLVP AEGED LKK QLM
Sbjct: 364 EQNRGKPNWEHLNEDLHVLITVEDAQNRAEIKLKRAVEEVKKLLVPAAEGEDSLKKMQLM 543

Query: 251 ELAIINGTYRDTTAKSVAVCDEEWRRLVAASDSRLLTSTGLPGLAAQIRAPA-AAPLGAP 309
           ELAI+NGTYRD   KS A+          A +  R++T         A +R P  A P   P
Sbjct: 544 ELAILNGTYRDANIKSPALAFS--LAATAQAAPRIITGPAPVLPPAALRTPTPAGPTIMP 717

Query: 310 LILNPRMTVPTTAASILSAQAAPTAAFDQTG--HGMIFAPYDYANYAALAGNPLLTEYAD 367
           LI   +  V       + +     PTAA    G    G+I+ PY+Y  Y        +L    +
Sbjct: 718 LIRQIQTAV------MPNGTPHPTAAIVPPGPEAGLIYTPYEYP-YTLAPATSILEYPIE 876

Query: 368 HS--VGAIKQQRRLATNREHPYQR 389
            S  +GA+  + R     R HPYQR
Sbjct: 877 PSGVLGAVATKVRRHDMRVHPYQR
```